ÁLGEBRA LINEAR

2ª EDIÇÃO

KENNETH HOFFMAN

Associate Professor of Mathematics
Massachusetts Institute of Technology

RAY KUNZE

Associate Professor of Mathematics
Washington University
St. Louis, Mo.

Tradução de

RENATE WATANABE

Professora de Matemática da Universidade Mackenzie

RIO DE JANEIRO
SÃO PAULO

LIVROS TÉCNICOS E CIENTÍFICOS EDITORA

COPYRIGHT © 1979, Kenneth Hoffman e Ray Kunze

Proibida a reprodução, mesmo
parcial, e por qualquer processo,
sem autorização expressa dos
Autores e Editor.

1ª edição – 1971
Reimpressões – 1973 e 1976
2ª edição – 1979

CIP-Brasil. Catalogação-na-fonte
Sindicato Nacional dos Editores de Livros, RJ.

Hoffman, Kenneth.
H647a Álgebra linear / Kenneth Hoffman [e] Ray Kunze; tradução de Renate Watanabe. – 2. ed. – Rio de Janeiro: Livros Técnicos e Científicos, 1979.

Tradução de: Linear algebra
Apêndice
Bibliografia

1. Álgebra linear 1. Kunze, Ray I. Título

CDD – 512.5
79-0363 CDU – 512.8

ISBN 85-216-0062-3

Direitos desta edição reservados:
LIVROS TÉCNICOS E CIENTÍFICOS EDITORA S. A.
Av. Venezuela, 163 – Centro
20220 – Rio de Janeiro – RJ
1979
Impresso no Brasil

PREFÁCIO

Nosso propósito original ao escrever este livro foi fornecer um texto para o curso de graduação em Álgebra Linear no *Massachusetts Institute of Technology.* Esse curso se destinava ao terceiro ano dos que optassem por Matemática, embora três quartos dos estudantes que o freqüentavam se especializassem em outras disciplinas tecnológicas e científicas e variassem desde calouros até estudantes de pós-graduação. Essa descrição da audiência do M.I.T. para o texto permanece, em linhas gerais, correta até hoje. Os dez anos decorridos desde a primeira edição viram a proliferação de cursos de Álgebra Linear por todo o país e ofereceram a um dos Autores a oportunidade de lecionar essa matéria básica para diversos grupos na *Brandeis University, Washington University* (St. Louis) e *University of California* (Irvine).

Nosso objetivo principal ao revisar o livro *Álgebra linear* foi adaptá-lo para uma maior variedade de cursos. De um lado, estruturamos os capítulos, principalmente os mais difíceis, de tal modo que existissem, ao longo do caminho, vários pontos naturais de parada, possibilitando ao instrutor escolher de diversas maneiras os tópicos para um curso trimestral ou semestral. Por outro lado, aumentamos a quantidade da matéria para que o texto pudesse ser usado em um curso de Álgebra Linear mais amplo, de um ano, ou mesmo como um livro de referência para matemáticos.

As maiores alterações ocorreram no nosso tratamento de formas canônicas e espaços com produto interno. No Cap. 6 não mais começamos com a teoria espacial geral que fundamenta a teoria das formas canônicas. Tratamos inicialmente os valores característicos em relação com teoremas de triangulação e diagonalização e em seguida escalamos o caminho para a teoria geral.

Separamos o Cap. 8 em dois, de modo que os tópicos fundamentais sobre espaços com produto interno e diagonalização unitária fossem seguidos de um novo Cap. 9, que, por sua vez, trata de formas sesquilineares e das propriedades mais sofisticadas de operadores normais, incluindo operadores normais sobre espaços com produto interno.

Introduzimos também pequenas alterações e melhoramentos na primeira edição, porém a filosofia subjacente do texto permaneceu inalterada.

Concessão alguma foi feita ao fato de a maioria dos alunos não estar interessada primordialmente em Matemática, porque acreditamos que um curso de Matemática não deveria fornecer a estudantes de Ciências, Engenharia ou Ciências Sociais um amontoado de métodos, e sim proporcionar a eles uma compreensão dos conceitos matemáticos fundamentais.

Por outro lado, estávamos profundamente conscientes da grande variação de conhecimentos que os estudantes poderiam possuir e, em particular, do fato de terem eles tido muito pouca experiência com o raciocínio matemático abstrato. Por essa razão, evitamos a introdução de muitas idéias abstratas logo no início do livro. Como complemento incluímos um Apêndice, onde são apresentadas idéias básicas tais como conjunto, função e relação de equivalência. Achamos mais proveitoso não insistir nessas idéias independentemente, e sim aconselhar os estudantes a ler o Apêndice à medida que elas surgissem.

Em todo o livro incluímos uma grande diversidade de exemplos dos conceitos importantes que ocorrem. O estudo de tais exemplos é de fundamental importância e tende a minimizar o número de estudantes que conseguem repetir definições, teoremas e demonstrações em ordem lógica, sem apreender o significado dos conceitos abstratos. O livro contém também uma ampla gama de exercícios graduados (em torno de seiscentos), que variam desde aplicações rotineiras aos que desafiarão até os melhores alunos. Pretendemos que esses exercícios sejam parte importante do texto.

O Cap. 1 trata de sistemas de equações lineares e sua resolução por meio de operações elementares sobre linhas de matrizes. Tem sido nosso costume despender seis aulas nessa matéria, o que proporciona ao estudante um esboço das origens da Álgebra Linear e das técnicas de cálculo computacionais necessárias ao entendimento de exemplos das idéias mais abstratas ocorrentes nos capítulos

posteriores. O Cap. 2 discorre sobre espaços vetoriais, subespaços, bases e dimensão. O Cap. 3 trata das transformações lineares, sua álgebra, sua representação por matrizes, bem como de isomorfismo, funcionais lineares e espaços duais.

O Cap. 4 define a álgebra dos polinômios sobre um corpo, os ideais naquela álgebra e a decomposição de um polinômio em fatores primos. Nele também são tratadas as raízes, a fórmula de Taylor e a fórmula de interpolação de Lagrange. O Cap. 5 desenvolve determinantes de matrizes quadradas, sendo o determinante encarado como uma função n-linear alternada das linhas de uma matriz, e prossegue com o estudo das funções multilineares sobre módulos e o anel de Grassman. A matéria sobre módulos coloca o conceito de determinante em um contexto mais amplo e abrangente do que o normalmente encontrado em livros elementares.

Os Caps. 6 e 7 contêm uma discussão dos conceitos fundamentais para a análise de uma só transformação linear sobre um espaço vetorial de dimensão finita; a análise de valores característicos, transformações trianguláveis e diagonalizáveis; os conceitos de partes diagonalizáveis e nilpotentes de uma transformação mais geral e as formas racional e canônica de Jordan. Os teoremas da decomposição primária e cíclica desempenham um papel central, chegando-se a este último por meio do estudo de subespaços admissíveis. O Cap. 7 inclui a discussão de matrizes sobre um domínio de polinômios, o cálculo de fatores invariantes e divisores elementares de uma matriz e o desenvolvimento da forma canônica de Smith. O capítulo termina com uma discussão sobre operadores semi-simples, para completar a análise de um só operador.

O Cap. 8 trata, com algum detalhe, dos espaços de dimensão finita com produto interno. Desenvolve a geometria básica, relacionando o conceito de ortogonalização à idéia de "melhor aproximação de um vetor" e passando aos conceitos de projeção ortogonal de um vetor sobre um subespaço e o suplementar ortogonal de um subespaço. O capítulo ainda trata dos operadores unitários e culmina com a diagonalização de operadores normais e auto-adjuntos.

O Cap. 9 introduz as formas sesquilineares, relaciona-as com os operadores positivos e auto-adjuntos sobre um espaço com produto interno, prossegue com a teoria espectral de operadores normais e, em seguida, com resultados mais sofisticados a respeito de operadores normais sobre espaços reais ou complexos com produto interno.

O Cap. 10 discute as formas bilineares, ressaltando as formas canônicas para formas simétricas e anti-simétricas, bem como grupos que conservam formas não-degeneradas, principalmente os grupos ortogonal, unitário, pseudo-ortogonal e de Lorentz.

Acreditamos que qualquer curso que use este texto deva desenvolver completamente os Caps. 1, 2 e 3 com a possível exclusão das Seçs. 3.6 e 3.7, que tratam do bidual e da transposta de uma transformação linear. Os Caps. 4 e 5, sobre polinômios e determinantes, podem ser tratados com diversos graus de profundidade. Ideais de polinômios e as propriedades fundamentais dos determinantes podem ser apenas esboçados, sem maior prejuízo do fluxo de lógica do texto; no entanto, sentimo-nos inclinados a desenvolver cuidadosamente esses capítulos (salvo os resultados sobre módulos), porque o seu conteúdo ilustra muito bem as idéias básicas da Álgebra Linear. Um curso elementar pode então ser satisfatoriamente concluído com as quatro primeiras seções do Cap. 6, juntamente com o (novo) Cap. 8. Se as formas racionais e as de Jordan forem incluídas, então será necessário desenvolver mais extensivamente o Cap. 6.

Nosso reconhecimento permanece para com aqueles que contribuíram na primeira edição, principalmente os Profs. Harry Furstenberg, Louis Howard, Daniel Kan, Edward Thorp, Sra. Judith Bowers, Sra. Betty Ann (Sargent) Rose e Sra. Phyllis Ruby. Queremos acrescentar nossos agradecimentos aos muitos estudantes e colegas cujos comentários perspícuos nos levaram a esta revisão e à equipe da Prentice-Hall, pela paciente colaboração com dois autores envolvidos nas agonias da administração acadêmica. Por último, nossos agradecimentos à Sra. Sophia Koulouras pela sua habilidade e incansável esforço na datilografia do manuscrito revisado.

K.M.H. / R.A.K.

SUMÁRIO

CAPÍTULO 1. EQUAÇÕES LINEARES 1
1.1. Corpos Comutativos 1
1.2. Sistemas de Equações Lineares 4
1.3. Matrizes e Operações Elementares sobre Linhas 7
1.4. Matrizes Linha-Reduzidas à Forma em Escada 14
1.5. Multiplicação de Matrizes 20
1.6. Matrizes Inversíveis 27

CAPÍTULO 2. ESPAÇOS VETORIAIS 35
2.1. Espaços Vetoriais 35
2.2. Subespaços 43
2.3. Bases e Dimensão 50
2.4. Coordenadas 61
2.5. Resumo de Linha-equivalência 70
2.6. Cálculos Concernentes a Subespaços 74

CAPÍTULO 3. TRANSFORMAÇÕES LINEARES 85
3.1. Transformações Lineares 85
3.2. A Álgebra das Transformações Lineares 94
3.3 Isomorfismo 107
3.4. Representação de Transformações por Matrizes 109
3.5. Funcionais Lineares 123
3.6. O Bidual 136
3.7. A Transposta de uma Transformação Linear 142

CAPÍTULO 4. POLINÔMIOS 148
4.1. Álgebras 148
4.2. A Algebra dos Polinômios 149
4.3. Interpolação de Lagrange 157
4.4. Ideais de Polinômios 161
4.5. A Decomposição de um Polinômio em Fatores Primos 170

CAPÍTULO 5. DETERMINANTES 178
5.1. Anéis Comutativos 178
5.2. Funções Determinantes 179

5.3. Permutações e a Unicidade dos Determinantes 190
5.4. Propriedades Adicionais dos Determinantes 198
5.5. Módulos 208
5.6. Funções Multilineares 210
5.7. O Anel de Grassman 221

CAPÍTULO 6. FORMAS CANÔNICAS ELEMENTARES 232
6.1. Introdução 232
6.2. Valores Característicos 233
6.3. Polinômios Anuladores 244
6.4. Subespaços Invariantes 254
6.5. Triangulação Simultânea; Diagonalização Simultânea 264
6.6. Decomposições em Soma Direta 267
6.7. Somas Diretas Invariantes 273
6.8. O Teorema da Decomposição Primária 280

CAPÍTULO 7. AS FORMAS RACIONAL E DE JORDAN 290
7.1. Subespaços Cíclicos e Anuladores 290
7.2. Decomposições Cíclicas e a Forma Racional 295
7.3. A Forma de Jordan 311
7.4. Cálculo dos Fatores Invariantes 321
7.5. Resumo; Operadores Semi-Simples 335

CAPÍTULO 8. ESPAÇOS COM PRODUTO INTERNO 345
8.1. Produtos Internos 345
8.2. Espaços com Produto Interno 354
8.3. Funcionais Lineares e Adjuntos 371
8.4. Operadores Unitários 383
8.5. Operadores Normais 398

CAPÍTULO 9. OPERADORES SOBRE ESPAÇOS COM PRODUTO INTERNO 406
9.1. Introdução 406
9.2. Formas sobre Espaços com Produto Interno 406
9.3. Formas Positivas 413
9.4. Mais sobre Formas 422
9.5. Teoria Espectral 427
9.6. Outras Propriedades dos Operadores Normais 444

CAPÍTULO 10. FORMAS BILINEARES 457
10.1. Formas Bilineares 457
10.2. Formas Bilineares Simétricas 468
10.3. Formas Bilineares Anti-Simétricas 478
10.4. Grupos que Conservam Formas Bilineares 483

APÊNDICE 491
A.1. Conjuntos 492
A.2. Funções 493
A.3. Relações de Equivalência 498
A.4. Espaços Quocientes 501

A.5. Relações de Equivalência em Álgebra Linear 506
A.6. O Axioma da Escolha 508
BIBLIOGRAFIA 510
ÍNDICE 511

Capítulo 1
EQUAÇÕES LINEARES

1.1 Corpos Comutativos

Supomos que o leitor tenha familiaridade com a álgebra elementar dos números reais e complexos. Para uma grande parte deste livro as propriedades algébricas dos números que usaremos podem ser facilmente deduzidas da pequena lista abaixo de propriedades da adição e da multiplicação. Indicamos por F o conjunto dos números reais ou o conjunto dos números complexos.

1. A adição é comutativa,

$$x + y = y + x$$

para todos x e y em F.

2. A adição é associativa,

$$x + (y + z) = (x + y) + z$$

para todos x, y e z em F.

3. Existe um único elemento 0 (zero) em F tal que $x + 0 = x$, para todo x em F.

4. A cada x em F corresponde um único elemento $(-x)$ em F tal que $x + (-x) = 0$.

5. A multiplicação é comutativa,

$$xy = yx$$

para todos x e y em F.

6. A multiplicação é associativa,

$$x(yz) = (xy)z$$

para todos x, y e z em F.

7. Existe um único elemento não-nulo 1 (um) em F tal que $x1 = x$, para todo x em F.

8. A cada x não-nulo em F corresponde um único x^{-1} (ou $1/x$) em F tal que $xx^{-1} = 1$.

9. A multiplicação é distributiva em relação à adição; isto é, $x(y + z) = xy + xz$, para todos x, y e z em F.

Suponhamos que se tenha um conjunto F de objetos $x, y, z, \ldots$ e duas operações sobre os elementos de F como segue. A primeira operação, denominada adição, associa a cada par de elementos x, y em F um elemento $(x + y)$ em F; a segunda operação, denominada multiplicação, associa a cada par x, y um elemento xy em F; e estas duas operações satisfazem as condições (1)-(9) acima. O conjunto F, munido destas duas operações, é então denominado um **corpo comutativo***. A grosso modo, um corpo é um conjunto munido de algumas operações sobre seus objetos, as quais se comportam como a adição, subtração, multiplicação e divisão usuais de números no sentido de que elas obedecem às nove regras de álgebra acima relacionadas. Com as propriedades usuais da adição e multiplicação, o conjunto C dos números complexos é um corpo, como o é o conjunto R dos números reais.

Na maior parte deste livro, os "números" que usamos podem ser os elementos de qualquer corpo F. Para permitir esta generalização, usaremos a palavra "escalar" ao invés de "número". O leitor não perderá muito se supuser sempre que o corpo de escalares seja um subcorpo do corpo dos números complexos. Um **subcorpo** do corpo C é um conjunto F de números complexos que é um corpo em relação às operações usuais de adição e multiplicação de números complexos. Isto significa que 0 e 1 estão no conjunto F e que se x e y são elementos de F então $(x + y)$, $-x$, xy e x^{-1} (se $x \neq 0$) também o são. Um exemplo de um subcorpo desta natureza é o corpo R dos números reais; de fato, se identificarmos os números reais com os números complexos $(a + ib)$ para os quais $b = 0$, o 0 e o 1 do corpo complexo são números reais e, se x e y são reais, $(x + y)$, $-x$, xy, e x^{-1} (se $x \neq 0$) também o são. Daremos outros

(*) Neste livro, sempre teremos corpos comutativos, portanto abreviaremos a denominação escrevendo simplesmente corpos (N. do T.)

exemplos abaixo. O objetivo de nossa discussão sobre subcorpos é essencialmente o seguinte: quando trabalhamos com escalares de um certo subcorpo de C, a realização das operações de adição, subtração, multiplicação ou divisão sobre estes escalares não nos tira daquele subcorpo.

Exemplo 1. O conjunto dos **inteiros positivos:** 1, 2, 3, ..., não é um subcorpo de C, por diversas razões. Por exemplo, 0 não é um inteiro positivo; para qualquer inteiro positivo n, $-n$ não é um inteiro positivo; para qualquer inteiro n, exceto 1, $1/n$ não é um inteiro positivo.

Exemplo 2. O conjunto dos **inteiros**: ..., −2, −1, 0, 1, 2, ..., não é um subcorpo de C, pois para um inteiro n, $1/n$ não é um inteiro a menos que n seja 1 ou -1. Com as operações usuais de adição e multiplicação, o conjunto dos inteiros satisfaz todas as condições (1)-(9) com exceção da condição (8).

Exemplo 3. O conjunto dos **números racionais,** isto é, números da forma p/q, onde p e q são inteiros e $q \neq 0$, é um subcorpo do corpo dos números complexos. A divisão, que não é possível dentro do conjunto dos inteiros, pode ser feita dentro do conjunto dos números racionais. O leitor interessado deve verificar que qualquer subcorpo de C contém todos os números racionais.

Exemplo 4. O conjunto de todos os números complexos da forma $x + y\sqrt{2}$, onde x e y são racionais, é um subcorpo de C. Deixamos a cargo do leitor a verificação deste fato.

Nos exemplos e exercícios deste livro, o leitor deve supor que o corpo em questão é um subcorpo dos números complexos, a não ser que seja explicitamente declarado que o corpo é mais geral. Não queremos nos estender sobre esse ponto mas, devemos indicar por que adotamos tal convenção. Se F é um corpo, pode ser possível adicionar um número finito de parcelas iguais à unidade 1 e obter 0 (veja o exercício 5 após a seção 1.2):

$$1 + 1 + \dots + 1 = 0.$$

Isto não acontece no corpo dos números complexos (ou em qualquer de seus subcorpos). Quando este fato acontecer em F, o menor n, tal que a soma de n 1's é 0, é chamado característica do

corpo F. Se este fato não acontecer em F, então (por alguma estranha razão) F se diz um corpo de característica zero. Freqüentemente quando supomos que F é um subcorpo de C, queremos é garantir que F seja um corpo de característica zero; mas, num primeiro contato com álgebra linear, é melhor, em geral, não nos preocuparmos muito com a característica de corpos.

1.2 Sistemas de Equações Lineares

Suponhamos que F seja um corpo. Consideremos o problema da determinação de n escalares (elementos de F) $x_1, \ldots, x_n$ que satisfaçam as condições

$$\begin{array}{ccccccc} A_{11}x_1 & + & A_{12}x_2 & + \ldots + & A_{1n}x_n & = & y_1 \\ A_{21}x_1 & + & A_{22}x_2 & + \ldots + & A_{2n}x_n & = & y_2 \\ \vdots & & \vdots & & \vdots & & \vdots \\ A_{m1}x_1 & + & A_{m2}x_2 & + \ldots + & A_{mn}x_n & = & y_m \end{array} \tag{1-1}$$

onde $y_1, \ldots, y_m$ e A_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, são elementos dados de F. Denominamos (1-1) um **sistema de m equações lineares a n incógnitas.** Toda n-upla $((x_1, \ldots, x_n)$ de elementos de F que satisfaz a cada uma das equações em (1-1) é dita uma **solução** do sistema. Se $y_1 = y_2 = \ldots = y_m = 0$, dizemos que o sistema é **homogêneo,** ou que cada uma das equações é homogênea.

O método mais importante para determinar as soluções de um sistema de equações lineares é talvez o método de eliminação. Podemos ilustrar este método com o sistema homogêneo

$$\begin{aligned} 2x_1 - x_2 + x_3 &= 0 \\ x_1 + 3x_2 + 4x_3 &= 0. \end{aligned}$$

Somando (-2) vezes a segunda equação à primeira equação obtemos

$$-7x_2 - 7x_3 = 0$$

ou $x_2 = -x_3$. Somando 3 vezes a primeira equação à segunda equação, obtemos

$$7x_1 + 7x_3 = 0$$

ou $x_1 = -x_3$. Assim, concluímos que se (x_1, x_2, x_3) é uma solução então $x_1 = x_2 = -x_3$. Reciprocamente, pode-se verificar pronta-

mente que toda terna deste tipo é uma solução. Assim, o conjunto de soluções consiste de todas as ternas $(-a, -a, a)$.

Determinamos as soluções deste sistema de equações "eliminando incógnitas", isto é, multiplicando equações por escalares e daí somando-as para obter equações em que alguns dos x_j não estejam presentes. Queremos formalizar ligeiramente este processo para que possamos compreender por que ele funciona e para que possamos efetuar os cálculos necessários para resolvermos um sistema de uma maneira organizada.

Para o sistema arbitrário (1-1), suponhamos selecionar m escalares, multiplicar a j-ésima equação por c_j e daí somar. Obtemos a equação

$$(c_1A_{11} + \dots + c_mA_{m1})\,x_1 + \dots + (c_1A_{1n} + \dots + c_mA_{mn})\,x_n = \\ = c_1y_1 + \dots + c_my_m.$$

Tal equação será por nós denominada uma **combinação linear** das equações em (1-1). Evidentemente, toda solução do sistema de equações (1-1) também será uma solução desta nova equação. Esta é a idéia fundamental do processo de eliminação. Se temos outro sistema de equações lineares

$$\begin{array}{ccc} B_{11}x_1 + \dots + B_{1n}x_n = z_1 \\ \vdots \\ B_{k1}x_1 + \dots + B_{kn}x_n = z_k \end{array} \tag{1-2}$$

no qual cada uma das k equações é uma combinação linear das equações em (1-1), então toda solução de (1-1) é uma solução deste novo sistema. É claro que pode acontecer que algumas soluções de (1-2) não sejam soluções de (1-1). Isto obviamente não acontece se cada equação do sistema original é uma combinação linear das equações do novo sistema. Diremos que dois sistemas de equações lineares são **equivalentes** se cada equação de cada sistema for uma combinação linear das equações do outro sistema. Podemos então enunciar formalmente nossas observações como segue.

Teorema 1. *Sistemas equivalentes de equações lineares têm exatamente as mesmas soluções.*

Para o processo de eliminação ser eficiente na determinação das soluções de um sistema como (1-1), é necessário que se saiba,

formando combinações lineares das equações dadas, como produzir um sistema equivalente de equações que seja mais fácil de resolver. Na próxima seção discutiremos um método para conseguir isto.

Exercícios

1. Verificar que o conjunto de números complexos descrito no Exemplo 4 é um subcorpo de C.

2. Seja F o corpo dos números complexos. Os dois seguintes sistemas de equações lineares são equivalentes? Em caso afirmativo, exprimir cada equação de cada sistema como uma combinação linear das equações do outro sistema.

$$\begin{aligned} x_1 - x_2 &= 0 \\ 2x_1 + x_2 &= 0 \end{aligned} \qquad \begin{aligned} 3x_1 + x_2 &= 0 \\ x_1 + x_2 &= 0 \end{aligned}$$

3. Repetir o Exercício 2 para os seguintes sistemas de equações:

$$\begin{aligned} -x_1 + x_2 + 4x_3 &= 0 \\ x_1 + 3x_2 + 8x_3 &= 0 \\ \tfrac{1}{2}x_1 + x_2 + \tfrac{5}{2}x_3 &= 0 \end{aligned} \qquad \begin{aligned} x_1 \qquad - x_3 &= 0 \\ x_2 + 3x_3 &= 0 \end{aligned}$$

4. Repetir o Exercício 2 para os sistemas seguintes:

$$\begin{aligned} 2x_1 + (-1 + i)x_2 \qquad + x_4 &= 0 \\ 3x_2 - 2ix_3 + 5x_4 &= 0 \end{aligned} \qquad \begin{aligned} \left(1 + \frac{i}{2}\right)x_1 + 8x_2 - ix_3 - x_4 &= 0 \\ \tfrac{2}{3}x_1 - \tfrac{1}{2}x_2 + x_3 + 7x_4 &= 0 \end{aligned}$$

5. Seja F um conjunto que contém exatamente dois elementos, 0 e 1. Definamos uma adição e uma multiplicação pelas tábuas:

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

Verificar que o conjunto F, munido destas duas operações, é um corpo.

6. Demonstrar que se dois sistemas homogêneos de equações lineares a duas incógnitas têm as mesmas soluções, então eles são equivalentes.

7. Demonstrar que todo subcorpo do corpo dos números complexos contém todos os números racionais.

8. Demonstrar que todo corpo de característica zero contém uma cópia do corpo dos números racionais.

1.3 Matrizes e Operações Elementares sobre Linhas

Não podemos deixar de observar que, ao formarmos combinações lineares de equações lineares, não há necessidade de continuarmos escrevendo as "incógnitas" $x_1, \ldots, x_n$, uma vez que, na realidade, fazemos cálculos apenas com os coeficientes A_{ij} e os escalares y_i. Abreviaremos o sistema (1-1) por

$$AX = Y$$

onde

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{e} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

Denominamos A a **matriz dos coeficientes** do sistema. Rigorosamente falando, a tabela retangular acima exibida não é uma matriz, mas sim uma representação de uma matriz. Uma $m \times n$ **matriz sobre o corpo** F é uma função A do conjunto dos pares de inteiros (i, j), $1 \leq i \leq m$, $1 \leq j \leq n$, no corpo F. Os **elementos** da matriz A são os escalares $A(i, j) = A_{ij}$ e, com bastante freqüência, o mais conveniente é descrever a matriz exibindo seus elementos numa tabela retangular com m linhas e n colunas, como acima. Assim X (acima) é, ou define uma $n \times 1$ matriz e Y é uma $m \times 1$ matriz. Por ora, $AX = Y$ nada mais é que uma notação taquigráfica para o nosso sistema de equações lineares. Posteriormente, quando houvermos definido uma multiplicação de matrizes, aquilo significará que Y é o produto de A por X.

Queremos agora considerar operações sobre linhas da matriz A que correspondam a formar combinações lineares das equações do sistema $AX = Y$. Restringiremos nossa atenção a três **operações elementares sobre as linhas** de uma $m \times n$ matriz A sobre o corpo F:

1. multiplicação de uma linha de A por um escalar c não-nulo;
2. substituição da r-ésima linha de A pela linha r mais c vezes a linha s, sendo c um escalar arbitrário e $r \neq s$;
3. transposição de duas linhas de A.

Uma operação elementar sobre linhas é assim um tipo particular de função (regra) e que associa a cada $m \times n$ matriz A uma $m \times n$ matriz $e(A)$. Pode-se descrever e com precisão nos três casos acima como segue:

1. $e(A)_{ij} = A_{ij}$ se $i \neq r$, $e(A)_{rj} = cA_{rj}$.
2. $e(A)_{ij} = A_{ij}$ se $i \neq r$, $e(A)_{rj} = A_{rj} + cA_{sj}$.
3. $e(A)_{ij} = A_{ij}$ se i é diferente de r e de s, $e(A)_{rj} = A_{sj}$, $e(A)_{sj} = A_{rj}$.

Ao definirmos $e(A)$ não importa muito o número de colunas de A, mas o número de linhas de A é crucial. Por exemplo, deve-se tomar cuidado ao decidir o que significa trocar as linhas 5 e 6 de uma 5×5 matriz. Para evitar tais complicações, convencionaremos que uma operação elementar e sobre as linhas é definida sobre a classe das $m \times n$ matrizes sobre F, para um certo m fixo mas para n arbitrário. Em outras palavras, um e particular é definido sobre a classe das matrizes com m linhas sobre F.

Uma razão para nos restringirmos a estes três tipos simples de operações sobre linhas é que, tendo efetuado uma tal operação e sobre uma matriz A, podemos voltar a A efetuando uma operação do mesmo tipo sobre $e(A)$.

Teorema 2. *A cada operação elementar sobre linhas* e *corresponde uma operação elementar sobre linhas* e₁ *tal que, do mesmo tipo que* e, $e_1(e(A)) = e(e_1(A)) = A$ *para qualquer* A. *Em outras palavras, a operação (função) inversa de uma operação elementar sobre linhas existe e é uma operação elementar sobre linhas do mesmo tipo.*

Demonstração. (1) Suponhamos que e seja a operação que multiplica a r-ésima linha de uma matriz pelo escalar não-nulo c. Seja e_1 a operação que multiplica a linha r por c^{-1}. (2) Suponhamos que e seja a operação que substitui a linha r pela linha r mais c vezes a linha s, $r \neq s$. Seja e_1 a operação que substitui a linha r pela linha r mais $(-c)$ vezes a linha s. (3) Se e transpõe as linhas r e s, seja $e_1 = e$. Em cada um destes três casos temos evidentemente $e_1(e(A)) = e(e_1(A)) = A$ para cada A.

Definição. *Se* A *e* B *são* $m \times n$ *matrizes sobre o corpo* F, *dizemos que* B **é linha-equivalente a** A *se* B *pode ser obtida de* A *por uma seqüência finita de operações elementares sobre linhas.*

Usando o Teorema 2, o leitor deverá achar fácil verificar o que segue. Toda matriz é linha-equivalente a si mesma; se B é linha-equivalente a A, então, A é linha-equivalente a B; se B é linha-equivalente a A e C é linha-equivalente a B, então C é linha-equivalente a A. Em outras palavras, a linha-equivalência é uma relação de equivalência (ver Apêndice).

Teorema 3. *Se* A *e* B *são* $m \times n$ *matrizes linha-equivalentes, os sistemas homogêneos de equações lineares* $AX = 0$ *e* $BX = 0$ *têm exatamente as mesmas soluções.*

Demonstração. Suponhamos passar de A para B por meio de uma seqüência finita de operações elementares sobre linhas:

$$A = A_0 \rightarrow A_2 \rightarrow \ldots \rightarrow A_k = B.$$

Basta demonstrar que os sistemas $A_jX = 0$ e $A_{j+1}X = 0$ têm as mesmas soluções, isto é, que uma operação elementar sobre linhas não altera o conjunto das soluções.

Assim, suponhamos que B seja obtida de A por uma única operação elementar sobre linhas. Qualquer que seja o tipo da operação, (1), (2) ou (3), cada equação dos sistemas $BX = 0$ será uma combinação linear das equações do sistema $AX = 0$. Como a inversa de uma operação elementar sobre linhas é uma operação elementar sobre linhas, cada equação em $AX = 0$ também será uma combinação linear das equações em $BX = 0$. Logo estes dois sistemas são equivalentes e, pelo Teorema 1, têm as mesmas soluções.

Exemplo 5. Suponhamos que F seja o corpo dos números racionais e que

$$A = \begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix}.$$

Efetuaremos uma seqüência finita de operações elementares sobre as linhas de A, indicando por números entre parênteses o tipo de operação efetuada.

$$\begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & -2 & -1 & 7 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & \frac{15}{2} & -\frac{55}{2} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & -\frac{5}{3} \end{bmatrix}$$

A linha-equivalência de A com a matriz final na seqüência acima nos diz em particular que as soluções de

$$\begin{aligned} 2x_1 - x_2 + 3x_3 + 2x_4 &= 0 \\ x_1 + 4x_2 \qquad\quad - x_4 &= 0 \\ 2x_1 + 6x_2 - x_3 + 5x_4 &= 0 \end{aligned}$$

e

$$\begin{aligned} x_3 - \tfrac{11}{3} x_4 &= 0 \\ x_1 + \tfrac{17}{3} x_4 &= 0 \\ x_2 - \tfrac{5}{3} x_4 &= 0 \end{aligned}$$

são exatamente as mesmas. No segundo sistema é evidente que atribuindo um valor racional arbitrário c a x_4, obtemos uma solução $(-\frac{17}{3}c, \frac{5}{3}c, \frac{11}{3}c, c)$, e também que toda solução é desta forma.

Exemplo 6. Suponhamos que F seja o corpo dos números complexos e que

$$A = \begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix}.$$

Ao efetuarmos operações sobre linhas freqüentemente convém combinar várias operações do tipo (2). Com isto em mente

$$\begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 2+i \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & 1 \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Assim o sistema de equações

$$\begin{aligned} -x_1 + ix_2 &= 0 \\ -ix_1 + 3x_2 &= 0 \\ ix_1 + 2x_2 &= 0 \end{aligned}$$

possui apenas a solução trivial $x_1 = x_2 = 0$.

Nos Exemplos 5 e 6 é óbvio que não efetuamos operações sobre linhas ao acaso. Nossa escolha de operações sobre linhas foi motivada por um desejo de simplificar a matriz dos coeficientes de uma maneira análoga à "eliminação de incógnitas" no sistema de equações lineares. Coloquemos agora uma definição formal do tipo da matriz à qual estávamos tentando chegar.

Definição. *Uma* m × n *matriz* R *é dita* **linha-reduzida** *se:*

(a) *o primeiro elemento não-nulo em cada linha não-nula de* R *é igual a* 1;

(b) *cada coluna de* R *que contém o primeiro elemento não-nulo de alguma linha tem todos os seus outros elementos nulos.*

Exemplo 7. Um exemplo de uma matriz linha-reduzida é a $n \times n$ **matriz** (quadrada) **unidade** I. Esta é a $n \times n$ matriz definida por

$$I_{ij} = \delta_{ij} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j. \end{cases}$$

Esta é a primeira de muitas ocasiões em que usaremos o **símbolo de Kronecker** (δ).

Nos Exemplos 5 e 6, as matrizes finais nas seqüências apresentadas são matrizes linha-reduzidas. Dois exemplos de matrizes que *não* são linha-reduzidas são:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

A segunda matriz não satisfaz a condição (a) porque o primeiro elemento não-nulo da primeira linha não é 1. A primeira matriz satisfaz a condição (a) mas não satisfaz a condição (b) na coluna 3.

Demonstraremos agora que podemos passar de uma matriz arbitrária a uma matriz linha-reduzida por meio de um número finito de operações elementares sobre linhas. Combinado com o Teorema 3, isto nos fornecerá um instrumento eficiente para a resolução de sistemas de equações lineares.

Teorema 4. *Toda* m × n *matriz sobre o corpo* F *é linha-equivalente a uma matriz linha-reduzida.*

Demonstração. Seja A uma $m \times n$ matriz sobre F. Se todo elemento na primeira linha de A é 0, então a condição (a) está satisfeita no que diz respeito à linha 1. Se a linha 1 tem um elemento não-nulo, seja k o menor inteiro positivo j para o qual $A_{1j} \neq 0$. Multipliquemos a linha 1 por A_{1k}^{-1} e então a condição (a) está satisfeita em relação à linha 1. Agora, para cada $i \geq 2$, somemos $(-A_{ik})$ vezes a linha 1 à linha i. Agora o primeiro elemento não-nulo da linha 1 ocorre na coluna k, este elemento é 1, e todos os outros elementos na coluna k são nulos.

Consideremos agora a matriz que resultou das operações acima. Se todo elemento na linha 2 é nulo, nada fazemos à linha 2. Se algum elemento na linha 2 é diferente de 0, multiplicamos a linha 2 por um escalar de modo que o primeiro elemento não-nulo seja 1. No caso em que a linha 1 tenha um primeiro elemento não-nulo; na coluna k, este primeiro elemento não-nulo na linha 2 não pode ocorrer na coluna k; digamos que ele aparece na coluna $k' \neq k$. Somando múltiplos adequados da linha 2 às diversas linhas, podemos fazer com que todos os elementos na coluna k' sejam nulos, com exceção do 1 na linha 2. O fato importante a ser notado é este: ao efetuarmos estas últimas operações, não alteramos os elementos da linha 1 nas colunas 1, ..., k; além disso, não alteramos nenhum elemento da coluna k. É claro que, se a linha 1 fosse identicamente nula, as operações com a linha 2 não afetariam a linha 1.

Trabalhando com uma linha de cada vez da maneira acima, é evidente que, com um número finito de passos, chegaremos a uma matriz linha-reduzida.

Exercícios

1. Determinar todas as soluções do sistema de equações

$$\begin{aligned}(1 - i)x_1 - ix_2 &= 0\\ 2x_1 + (1 - i)x_2 &= 0.\end{aligned}$$

2. Se

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}$$

determinar todas as soluções de $AX = 0$, tornando A linha-reduzida.

3. Se

$$A = \begin{bmatrix} 6 & -4 & 0 \\ 4 & -2 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

determinar todas as soluções de $AX = 2X$ e todas as soluções de $AX = 3X$. (O símbolo cX indica a matriz cujos elementos são c vezes os elementos correspondentes de X.)

4. Encontrar uma matriz linha-reduzida que seja linha-equivalente a

$$A = \begin{bmatrix} i & -(1+i) & 0 \\ 1 & -2 & 1 \\ 1 & 2i & -1 \end{bmatrix}.$$

5. Demonstrar que as duas matrizes seguintes *não* são linha-equivalentes:

$$\begin{bmatrix} 2 & 0 & 0 \\ a & -1 & 0 \\ b & c & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 2 \\ -2 & 0 & -1 \\ 1 & 3 & 5 \end{bmatrix}.$$

6. Seja

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

uma 2×2 matriz com elementos complexos. Suponhamos que A seja linha-reduzida e também que $a + b + c + d = 0$. Demonstrar que existem exatamente três destas matrizes.

7. Demonstrar que a transposição de duas linhas de uma matriz pode ser conseguida por uma seqüência finita de operações elementares sobre linhas dos outros dois tipos.

8. Consideremos o sistema de equações $AX = 0$ onde

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

é uma 2×2 matriz sobre o corpo F. Demonstrar o que segue.

(*a*) Se todo elemento de A é nulo, então todo par (x_1, x_2) é uma solução de $AX = 0$.

(*b*) Se $ad - bc \neq 0$, o sistema $AX = 0$ possui apenas a solução trivial $x_1 = x_2 = 0$.

(*c*) Se $ad - bc = 0$ e algum elemento de A é diferente de 0, então existe uma solução (x_1^0, x_2^0) tal que (x_1, x_2) é uma solução se, e somente se, existe um escalar y tal que $x_1 = yx_1^0$, $x_2 = yx_2^0$.

1.4 Matrizes Linha-reduzidas à Forma em Escada

Até agora, nosso trabalho com sistemas de equações lineares foi motivado por uma tentativa de determinar as soluções de um tal sistema. Na Seção 1.3 estabelecemos um método padronizado para determinar estas soluções. Desejamos agora obter algum conhecimento que seja um pouco mais teórico, e para tal propósito é conveniente ir um pouco além de matrizes linha-reduzidas.

Definição. *Uma* $m \times n$ *matriz* R *é dita uma* **matriz linha-reduzida à forma em escada** *se*

(a) R *é linha-reduzida*;

(b) *toda linha de* R *cujos elementos são todos nulos ocorre abaixo de todas as linhas que possuem um elemento não-nulo*;

(c) *se as linhas* 1, ..., r *são as linhas não-nulas de* R *e se o primeiro elemento não-nulo da linha* i *ocorre na coluna* k_i, $i = 1, \ldots, r$, *então* $k_1 < k_2 < \ldots < k_r$.

Pode-se também descrever uma $m \times n$ matriz R linha-reduzida à forma em escada como segue. Todo elemento em R é nulo ou então existe um inteiro positivo r, $1 \leq r \leq m$, e r inteiros positivos $k_1, \ldots, k_r$ com $1 \leq k_i \leq n$ e

(a) $R_{ij} = 0$ para $i > r$, e $R_{ij} = 0$ se $j < k_i$.

(b) $R_{ikj} = \delta_{ij}$, $1 \leq i \leq r$, $1 \leq j \leq r$.

(c) $k_1 < \ldots < k_r$.

Exemplo 8. Dois exemplos de matrizes linha-reduzidas à forma em escada são $n \times n$ matriz unidade e a $m \times n$ matriz nula $0^{m,n}$, na qual todos os elementos são nulos. O leitor não deverá encontrar nenhuma dificuldade para encontrar outros exemplos, mas gostaríamos de dar mais um exemplo não-trivial:

$$\begin{bmatrix} 0 & 1 & -3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Teorema 5. *Toda* m × n *matriz* A *é linha-equivalente a uma matriz linha-reduzida à forma em escada.*

Demonstração. Sabemos que A é linha-equivalente a uma matriz linha-reduzida. Portanto, basta observar que, efetuando um número finito de permutações das linhas de uma matriz linha-reduzida, podemos transformá-la numa matriz linha-reduzida à forma em escada.

Nos Exemplos 5 e 6, vimos a importância de matrizes linha-reduzidas na solução de sistemas homogêneos de equações lineares. Discutamos rapidamente o sistema $RX = 0$, no caso em que R é uma matriz linha-reduzida à forma em escada. Sejam as linhas $1, \ldots, r$ as linhas não-nulas de R e suponhamos que o primeiro elemento não-nulo da linha i ocorra na coluna k_i. O sistema $RX = 0$ consiste então de r equações não-triviais. Além disso, a incógnita x_{k_i} aparecerá (com coeficiente não-nulo) apenas na *i-ésima* equação. Se indicarmos por $u_1, \ldots, u_{n-r}$ as $(n - r)$ incógnitas que são diferentes de $x_{k_1}, \ldots, x_{k_r}$, então as r equações não-triviais em $RX = 0$ são da forma

$$\text{(1-3)} \qquad \begin{aligned} x_{k_1} + \sum_{j=1}^{n-r} C_{1j} u_j &= 0 \\ \vdots \qquad & \quad \vdots \\ x_{k_r} + \sum_{j=1}^{n-r} C_{rj} u_j &= 0. \end{aligned}$$

Todas as soluções dos sistemas de equações $RX = 0$ são obtidas atribuindo-se valores arbitrários a $u_1, \ldots, u_{n-r}$ e calculando os valores correspondentes de $x_{k_1}, \ldots, x_{k_r}$ por meio de (1-3). Por exem-

plo, se R é a matriz do exemplo 8 acima, então $r = 2$, $k_1 = 2$, $k_2 = 4$, e as duas equações não-triviais do sistema $RX = 0$ são

$$\begin{aligned} x_2 - 3x_3 \qquad + \tfrac{1}{2}x_5 &= 0 \quad \text{ou} \quad x_2 = 3x_3 - \tfrac{1}{2}x_5 \\ x_4 + 2x_5 &= 0 \quad \text{ou} \quad x_4 = 2x_5. \end{aligned}$$

Assim, podemos atribuir valores arbitrários a x_1, x_3 e x_5, digamos $x_1 = a$, $x_3 = b$, $x_5 = c$, e obter a solução $(a, 3b - \frac{1}{2}c,\ b, -2c,\ c)$.

Observemos mais um fato sobre o sistema de equações $RX = 0$. Se o número r de linhas não-nulas de R é menor que n, então o sistema $RX = 0$ admite uma solução não-trivial, isto é, uma solução $(x_1, \ldots, x_n)$ em que nem todo x_j é nulo. De fato, como $r < n$, podemos tomar algum x_j que não esteja entre as r incógnitas $x_{k_1}, \ldots, x_{k_r}$ e daí construir uma solução como acima na qual este x_j é 1. Esta observação nos leva a um dos conceitos mais fundamentais relativos a sistemas de equações lineares homogêneas.

Teorema 6. *Se* A *é uma* m × n *matriz e* m < n, *então o sistema homogêneo de equações lineares* AX = 0 *admite uma solução não-trivial.*

Demonstração. Seja R uma matriz linha-reduzida à forma em escada que seja linha-equivalente a A. Então os sistemas $AX = 0$ e $RX = 0$ possuem, pelo Teorema 3, as mesmas soluções. Se r é o número de linhas não-nulas em R, então certamente $r \leq m$ e como $m < n$, temos $r < n$. Decorre imediatamente de nossas observações acima que $AX = 0$ admite uma solução não-trivial.

Teorema 7. *Se* A *é uma* n × n *matriz (quadrada), então* A *é linha equivalente a* n × n *matriz unidade se, e somente se, o sistema de equações* AX = 0 *possuir apenas a solução trivial.*

Demonstração. Se A é linha-equivalente a I, então $AX = 0$ e $IX = 0$ têm as mesmas soluções. Reciprocamente, suponhamos que $AX = 0$ admita somente a solução trivial $X = 0$. Seja R uma $n \times n$ matriz linha-reduzida à forma em escada que seja linha-equivalente a A, e seja r o número de linhas não nulas de R. Então $RX = 0$ não admite solução não-trivial. Assim, $r \geq n$. Mas como R possui n linhas, certamente $r \leq n$ e temos $r = n$. Como isto significa que R possui na verdade um primeiro elemento não-nulo igual a 1 em cada uma de suas n linhas e como estes 1 ocorrem cada um numa das n colunas, R é, necessariamente, a $n \times n$ matriz unidade.

Perguntemos agora que operações elementares sobre linhas efetuar para resolver um sistema de equações lineares $AX = Y$ que não seja homogêneo. De início, devemos observar uma diferença básica entre este caso e o caso homogêneo, a saber, que enquanto o sistema homogêneo sempre admite a solução trivial $x_1 = \dots = x_n = 0$, um sistema não-homogêneo pode não ter nenhuma solução.

Formemos a matriz completa A' do sistema $AX = Y$. Esta é a $m \times (n+1)$ matriz cujas n primeiras colunas são as colunas de A e cuja última coluna é Y. Mais precisamente

$$A'_{ij} = A_{ij}, \text{ se } j \leq n$$
$$A'_{i(n+1)} = y_i$$

Suponhamos que efetuemos uma seqüência de operações elementares sobre as linhas de A, obtendo uma matriz R linha-reduzida à forma em escada. Se efetuarmos esta mesma seqüência de operações sobre a matriz completa A', obteremos uma matriz R' cujas n primeiras colunas são as colunas de R e cuja última coluna contém certos escalares $z_1, \dots, z_m$. Os escalares z_i são os elementos da $m \times 1$ matriz

$$Z = \begin{bmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_m \end{bmatrix}$$

que resulta de se aplicar a seqüência de operações sobre as linhas da matriz Y. Deve ser evidente ao leitor que, como na demonstração do Teorema 3, os sistemas $AX = Y$ e $RX = Z$ são equivalentes e portanto admitem as mesmas soluções. É bem fácil saber se o sistema $RX = Z$ possui soluções e em caso afirmativo determinar todas as soluções. De fato, se R possuir r linhas não-nulas, com o primeiro elemento não-nulo da linha i ocorrendo na coluna k_i, $i = 1, \dots, r$, então as r primeiras equações de $RX = Z$ exprimem realmente $x_{k_1}, \dots, x_{k_r}$ em termos dos $(n - r)$ x_j restantes e dos escalares $z_1, \dots, z_r$. As $(m - r)$ últimas equações são

$$\begin{aligned} 0 &= z_{r+1} \\ \cdot \quad & \quad \cdot \\ \cdot \quad & \quad \cdot \\ \cdot \quad & \quad \cdot \\ 0 &= z_m \end{aligned}$$

portanto a condição para o sistema ter uma solução é que $z_i = 0$ para $i > r$. Se esta condição é satisfeita, todas as soluções deste sistema podem ser determinadas, como no caso homogêneo, atribuindo-se valores arbitrários a $(n - r)$ dos x_j e daí calculando x_{k_i} por meio da i-ésima equação.

Exemplo 9. Seja F o corpo dos números racionais e

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 0 & 5 & -1 \end{bmatrix}$$

e suponhamos que se deseje resolver o sistema $AX = Y$ para certos y_1, y_2 e y_3. Efetuemos uma seqüência de operações sobre as linhas da matriz completa A' que torne A linha-reduzida:

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 2 & 1 & 1 & y_2 \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & 0 & \frac{3}{5} & \frac{1}{5}(y_1 + 2y_2) \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix}.$$

A condição para que o sistema $AX = Y$ tenha uma solução é portanto

$$2y_1 - y_2 + y_3 = 0$$

e se os escalares y_i dados satisfazem esta condição, todas as soluções são obtidas atribuindo-se um valor c a x_3 e depois calculando

$$\begin{aligned} x_1 &= \tfrac{3}{5}c + \tfrac{1}{5}(y_1 + 2y_2) \\ x_2 &= \tfrac{1}{5}c + \tfrac{1}{5}(y_2 - 2y_1). \end{aligned}$$

Façamos uma observação final sobre o sistema $AX = Y$. Suponhamos que os elementos da matriz A e os escalares $y_1, \ldots, y_m$ estejam num subcorpo F_1 do corpo F. Se o sistema de equações $AX = Y$ admite uma solução com $x_1, \ldots, x_n$ em F, ele admite uma solução com $x_1, \ldots, x_n$ em F_1. De fato, sobre qualquer um dos dois corpos, a condição para o sistema admitir uma solução é que valham certas relações entre $y_1, \ldots, y_m$ em F_1 (a saber, as relações $z_i = 0$ para $i > r$, acima). Por exemplo, se $AX = Y$ é um sistema de equações lineares no qual os escalares y_k e A_{ij} são números reais e, se existe uma solução na qual $x_1, \ldots, x_n$ são números complexos, então existe uma solução com $x_1, \ldots, x_n$ números reais.

Exercícios

1. Determinar todas as soluções do seguinte sistema de equações linha-reduzindo a matriz dos coeficientes:

$$\begin{aligned} \tfrac{1}{3}x_1 + 2x_2 - 6x_3 &= 0 \\ -4x_1 \qquad\quad + 5x_3 &= 0 \\ -3x_1 + 6x_2 - 13x_3 &= 0 \\ -\tfrac{7}{3}x_1 + 2x_2 - \tfrac{8}{3}x_3 &= 0 \end{aligned}$$

2. Determinar uma matriz linha-reduzida à forma em escada que seja equivalente a

$$A = \begin{bmatrix} 1 & -i \\ 2 & 2 \\ i & 1+i \end{bmatrix}.$$

Quais são as soluções de $AX = 0$?

3. Descrever explicitamente todas as 2×2 matrizes linha-reduzidas à forma em escada.

4. Consideremos o sistema de equações

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 \qquad + 2x_3 &= 1 \\ x_1 - 3x_2 + 4x_3 &= 2. \end{aligned}$$

Este sistema admite solução? Em caso afirmativo, descrever explicitamente todas as soluções.

5. Dar um exemplo de um sistema de duas equações lineares a duas incógnitas que não admite solução.

6. Mostrar que o sistema

$$\begin{aligned} x_1 - 2x_2 + x_3 + 2x_4 &= 1 \\ x_1 + x_2 - x_3 + x_4 &= 2 \\ x_1 + 7x_2 - 5x_3 - x_4 &= 3 \end{aligned}$$

não admite solução.

7. Determinar todas as soluções de

$$\begin{aligned} 2x_1 - 3x_2 - 7x_3 + 5x_4 + 2x_5 &= -2 \\ x_1 - 2x_2 - 4x_3 + 3x_4 + x_5 &= -2 \\ 2x_1 - 4x_3 + 2x_4 + x_5 &= 3 \\ x_1 - 5x_2 - 7x_3 + 6x_4 + 2x_5 &= -7. \end{aligned}$$

8. Seja

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}.$$

Para que ternas (y_1, y_2, y_3) o sistema $AX = Y$ admite solução?

9. *Seja*

$$A = \begin{bmatrix} 3 & -6 & 2 & -1 \\ -2 & 4 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 1 & -2 & 1 & 0 \end{bmatrix}.$$

Para que (y_1, y_2, y_3, y_4) o sistema de equações $AX = Y$ admite solução?

10. Suponhamos que R e R', sejam 2×3 matrizes linha-reduzidas à forma em escada e que os sistemas $RX = 0$ e $R'X = 0$ admitam as mesmas soluções. Demonstrar que $R = R'$.

1.5 Multiplicação de Matrizes

É evidente (ou, de qualquer modo, deveria ser) que o processo de formar combinações lineares das linhas de uma matriz é um processo fundamental. Por esta razão é vantajoso introduzir um esquema sistemático para indicar exatamente que operações devem ser efetuadas. Mais especificamente, suponhamos que B seja uma $n \times p$ matriz sobre um corpo F com linhas $\beta_1, \ldots, \beta_n$ e que a partir de B construamos uma matriz C com linhas $\gamma_1, \ldots, \gamma_m$ formando certas combinações lineares

$$\gamma_i = A_{i1}\beta_1 + A_{i2}\beta_2 + \ldots + A_{in}\beta_n. \tag{1-4}$$

As linhas de C são determinadas pelos mn escalares A_{ij} que são os elementos de uma $m \times n$ matriz A. Se (1-4) é desenvolvido como

$$(C_{i1} \dots C_{ip}) = \sum_{r=1}^{n} (A_{ir}B_{r1} \dots A_{ir}B_{rp})$$

vemos que os elementos de C são dados por

$$C_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj}.$$

Definição. *Seja* A *uma* $m \times n$ *matriz sobre o corpo* F *e seja* B *uma* $n \times p$ *matriz sobre* F. *O* **produto** AB *é a* $m \times p$ *matriz* C *cujo elemento* ij *e*

$$C_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj}.$$

Exemplo 10. Eis alguns produtos de matrizes com elementos racionais.

(a)
$$\begin{bmatrix} 5 & -1 & 2 \\ 0 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 5 & -1 & 2 \\ 15 & 4 & 8 \end{bmatrix}$$

Neste caso

$$\begin{aligned} \gamma_1 &= (5 \;\; -1 \;\; 2) = 1 \cdot (5 \;\; -1 \;\; 2) + 0 \cdot (15 \;\; 4 \;\; 8) \\ \gamma_2 &= (0 \;\; 7 \;\; 2) = -3(5 \;\; -1 \;\; 2) + 1 \cdot (15 \;\; 4 \;\; 8) \end{aligned}$$

(b)
$$\begin{bmatrix} 0 & 6 & 1 \\ 9 & 12 & -8 \\ 12 & 62 & -3 \\ 3 & 8 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 3 \\ 5 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 1 \\ 3 & 8 & -2 \end{bmatrix}$$

Neste caso

$$\begin{aligned} \gamma_2 &= (9 \;\; 12 \;\; -8) = -2(0 \;\; 6 \;\; 1) + 3(3 \;\; 8 \;\; -2) \\ \gamma_3 &= (12 \;\; 62 \;\; -3) = 5(0 \;\; 6 \;\; 1) + 4(3 \;\; 8 \;\; -2) \end{aligned}$$

(c)
$$\begin{bmatrix} 8 \\ 29 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

(d)
$$\begin{bmatrix} -2 & -4 \\ 6 & 12 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix}$$

Neste caso

$$\gamma_2 = (6 \;\; 12) = 3\,(2 \;\; 4)$$

(e)
$$\begin{bmatrix} 2 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \end{bmatrix}$$

(f)
$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -5 & 2 \\ 2 & 3 & 4 \\ 9 & -1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(g)
$$\begin{bmatrix} 1 & -5 & 2 \\ 2 & 3 & 4 \\ 9 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 9 & 0 \end{bmatrix}$$

É importante observar que o produto de duas matrizes pode não estar definido; o produto é definido se, e somente se, o número de colunas da primeira matriz coincide com o número de linhas da segunda matriz. Assim, não faz sentido trocar a ordem dos fatores em (a), (b) e (c) acima. Freqüentemente escreveremos produtos como AB sem mencionar explicitamente as dimensões dos fatores e, em tais casos, estará subentendido que o produto está definido. De (d), (e), (f), (g) vemos que mesmo quando ambos os produtos AB e BA estão definidos não é necessariamente verdade que $AB = BA$; em outras palavras a multiplicação de matrizes *não é comutativa.*

Exemplo 11.

(a) Se I é a $m \times m$ matriz unidade e A é uma $m \times n$ matriz, $IA = A$.

(b) Se I é a $n \times n$ matriz unidade e A é uma $m \times n$ matriz, $AI = A$.

(c) Se $0^{k,m}$ é a $k \times m$ matriz nula, $0^{k,n} = 0^{k,m}A$. Analogamente, $A0^{n,p} = 0^{m,p}$.

Exemplo 12. Seja A uma $m \times n$ matriz sobre F. Nossa notação taquigráfica anterior, $AX = Y$, para sistemas de equações lineares, é coerente com nossa definição de produtos de matrizes. De fato, se

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

com x_i em F, então AX é a $m \times 1$ matriz

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_m \end{bmatrix}$$

tal que $y_i = A_{i1}x_1 + A_{i2}x_2 + \dots + A_{in}x_n$.

O uso de matrizes-coluna sugere uma notação que freqüentemente é útil. Se B for uma $n \times p$ matriz, as colunas de B são as $1 \times n$ matrizes $B_1, \dots, B_p$ definidas por

$n \times 1$

$$B_j = \begin{bmatrix} B_{1j} \\ \cdot \\ \cdot \\ \cdot \\ B_{nj} \end{bmatrix}, \quad 1 \le j \le p$$

A matriz B é a sucessão destas colunas:

$$B = [B_1, \dots, B_p].$$

O elemento i, j da matriz produto AB é formado a partir da i-ésima linha de A e a j-ésima coluna de B. O leitor deve verificar que a j-ésima coluna de AB é AB_j:

$$AB = [AB_1, \dots, AB_p].$$

A despeito do fato de que um produto de matrizes depende da ordem em que os fatores são escritos, ele é independente da maneira pela qual elas são associadas, como o próximo teorema mostra.

Teorema 8. *Se* A, B, C *são matrizes sobre o corpo F tais que os produtos* BC *e* A(BC) *são definidos, então estão definidos os produtos* AB, (AB)C *e*

$$A(BC) = (AB)C.$$

Demonstração. Suponhamos que B seja uma $n \times p$ matriz. Como BC está definida, C é uma matriz com p linhas e BC tem n linhas. Como $A(BC)$ está definido podemos supor que A é uma $m \times n$ matriz. Assim, o produto AB existe e é uma $m \times p$ matriz, do que segue que o produto $(AB)C$ existe. Mostrar que $A(BC) = (AB)C$ significa mostrar que

$$[A(BC)]_{ij} = [(AB)C]_{ij}$$

para cada i, j. Por definição,

$$\begin{aligned}
[A(BC)]_{ij} &= \sum_r A_{ir}(BC)_{rj} \\
&= \sum_r A_{ir} \sum_s B_{rs}C_{sj} \\
&= \sum_r \sum_s A_{ir}B_{rs}C_{sj} \\
&= \sum_s \sum_r A_{ir}B_{rs}C_{sj} \\
&= \sum_s \left(\sum_r A_{ir}B_{rs}\right) C_{sj} \\
&= \sum_s (AB)_{is}C_{sj} \\
&= [(AB)\,C]_{ij}.
\end{aligned}$$

Quando A é uma $n \times m$ matriz (quadrada), o produto AA está definido. Indicaremos esta matriz por A^2. Pelo Teorema 8, $(AA)A = A(AA)$ ou $A^2A = AA^2$, de modo que o produto AAA está definido sem ambigüidade. Indicaremos este produto por A^3. Em geral, o produto $AA \ldots A$ (k vezes) está definido sem ambigüidade e indicaremos este produto por A^k.

Notemos que a relação $A(BC) = (AB)C$ implica, entre outras coisas, que combinações lineares de combinações lineares das linhas de C são novamente combinações lineares das linhas de C.

Se B é uma dada matriz e C é obtida de B por meio de uma operação elementar sobre linhas, então cada linha de C é uma combinação linear das linhas de B, logo existe uma matriz A tal que $AB = C$. Em geral, existem muitas dessas matrizes A e, dentre elas todas, é conveniente e possível escolher uma que tenha um número de propriedades especiais. Antes de passar a isto precisamos introduzir uma classe de matrizes.

Definição. *Uma* m × m *matriz é dita uma* **matriz elementar** *se ela pode ser obtida da* m × m *matriz unidade por meio de uma única operação elementar sobre linhas.*

Exemplo 13. Uma 2 × 2 matriz elementar é necessariamente uma das seguintes:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$$

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}, \quad c \neq 0, \quad \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}, \quad c \neq 0.$$

Teorema 9. *Seja* e *uma operação elementar sobre linhas e seja* E *a* m × m *matriz elementar* E = e(I). *Então, para toda* m × n *matriz* A,

$$\mathrm{e(A) = EA.}$$

Demonstração. O ponto-chave da demonstração é que o elemento da i-ésima linha e j-ésima coluna da matriz produto EA é obtido da i-ésima linha de E e da j-ésima coluna de A. Os três tipos de operações elementares sobre linhas devem ser tratados separadamente. Daremos uma demonstração detalhada para uma operação do tipo (2). Os outros dois casos são de tratamento ainda mais simples que este, e serão deixados como exercícios. Suponhamos $r \neq s$ e seja e a operação "substituição da linha r pela linha r mais c vezes a linha s". Então,

$$E_{ik} = \begin{cases} \delta_{ik}, & i \neq r \\ \delta_{rk} + c\delta_{sk}, & i = r. \end{cases}$$

Portanto,

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik}A_{kj} = \begin{cases} A_{ik}, & i \neq r \\ A_{rj} + cA_{sj}, & i = r. \end{cases}$$

Em outras palavras, $EA = e(A)$.

Corolário. *Sejam* A *e* B m × n *matrizes sobre o corpo* F. *Então* B *é linha-equivalente a* A *se, e somente se,* B = PA, *onde* P *é um produto de* m × m *matrizes elementares.*

Demonstração. Suponhamos que $B = PA$, onde $P = E_s \ldots E_2E_1$ e as E_i são $m \times m$ matrizes elementares. Então E_1A é linha-equivalente a A e $E_2(E_1A)$ é linha-equivalente a E_1A. Assim E_2E_1A é linha-equivalente a A; continuando desta maneira, vemos que $(E_s \ldots E_1)A$ é linha-equivalente a A.

Suponhamos agora que B seja linha-equivalente a A. Sejam $E_1, E_2, \ldots, E_s$ as matrizes elementares correspondentes a alguma seqüência de operações elementares sobre linhas que levam A em B. Então $B = (E_s \ldots E_1)A$.

Exercícios

1. Sejam

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}, \quad C = [1 \quad -1].$$

Calcular ABC e CAB.

2. Sejam

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & 1 \\ 3 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -2 \\ 1 & 3 \\ 4 & 4 \end{bmatrix}.$$

Verificar diretamente que $A(AB) = A^2B$.

3. Determinar duas 2×2 matrizes A distintas tais que $A^2 = 0$ mas $A \neq 0$.

4. Para a matriz A do Exercício 2, determinar matrizes elementares $E_1, E_2, \ldots, E_k$ tais que

$$E_k, \ldots, E_2E_1A = I.$$

5. Sejam

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 2 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 1 \\ -4 & 4 \end{bmatrix}.$$

Existe alguma matriz C tal que $CA = B$?

6. Seja A uma $m \times n$ matriz e B uma $n \times k$ matriz. Mostrar que as colunas de $C = AB$ são combinações lineares das colunas de A. Se $\alpha_1, \ldots, \alpha_n$ são as colunas de A e $\gamma_1, \ldots, \gamma_k$ são as colunas de C, então

$$\gamma_j = \sum_{r=1}^{n} B_{rj}\alpha_r.$$

7. Sejam A e B duas 2×2 matrizes tais que $AB = I$. Demonstrar que $BA = I$.
8. Seja

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

uma 2×2 matriz. Perguntamos quando é possível encontrar 2×2 matrizes A e B tais que $C = AB - BA$. Demonstrar que tais matrizes podem ser encontradas se, e somente se, $C_{11} + C_{22} = 0$.

1.6 Matrizes Inversíveis

Suponhamos que P seja uma $m \times m$ matriz que seja um produto de matrizes elementares. Para cada $m \times n$ matriz A, a matriz $B = PA$ é linha-equivalente a A; logo A é linha-equivalente a B e existe um produto Q de matrizes elementares tal que $A = QB$. Em particular, isto é válido quando A é a $m \times m$ matriz unidade. Em outras palavras, existe uma $m \times m$ matriz Q, que é um produto de matrizes elementares, tal que $QP = I$. Como logo veremos, a existência de uma Q tal que $QP = I$ é equivalente ao fato de P ser um produto de matrizes elementares.

Definição. *Seja* A *uma* n × n *matriz* (*quadrada*) *sobre o corpo* F. *Uma* n × n *matriz* B *tal que* BA = I *é dita uma* **inversa à esquerda** *de* A; *uma* n × n *matriz* B *tal que* AB = I *é dita uma* **inversa à direita** *de* A. *Se* AB = BA = I, *então* A *é dita* **inversível.**

Lema. Se A possui uma inversa à esquerda B e uma inversa à direita C, então $B = C$.

Demonstração. Suponhamos que $BA = I$ e $AC = I$. Então

$$B = BI = B(AC) = (BA)C = IC = C.$$

Assim, se A possui uma inversa à esquerda e uma à direita, A é inversível e possui uma única inversa bilateral, que indicaremos por A^{-1} e denominaremos simplesmente a *inversa* de A.

Teorema 10. *Sejam* A *e* B $n \times n$ *matrizes sobre* F.

(i) *Se* A *é inversível,* A^{-1} *também o é e* $(A^{-1})^{-1} = A$.

(ii) *Se* A *e* B *são inversíveis,* AB *também o é e* $(AB)^{-1} = B^{-1}A^{-1}$.

Demonstração. A primeira afirmação é evidente pela simetria da definição. A segunda decorre da verificação das relações

$$(AB)(B^{-1}A^{-1}) = (B^{-1}A^{-1})(AB) = I.$$

Corolário. *Um produto de matrizes inversíveis é inversível.*

Teorema 11. *Uma matriz elementar é inversível.*

Demonstração. Seja E uma matriz elementar correspondente à operação elementar sobre linhas e. Se e_1 é a operação inversa de e (Teorema 2) e $E_1 = e_1(I)$, então

$$EE_1 = e(E_1) = e(e_1(I)) = I$$

$$E_1E = e_1(E) = e_1(e(I)) = I$$

de modo que E é inversível e $E_1 = E^{-1}$.

Exemplo 14.

(a) $$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(b) $$\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -c \\ 0 & 1 \end{bmatrix}$$

(c) $$\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix}$$

(d) Quando $c \neq 0$,

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} c^{-1} & 0 \\ 0 & 1 \end{bmatrix} \text{ e } \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & c^{-1} \end{bmatrix}.$$

Teorema 12. *Se* A *é uma* $n \times n$ *matriz, as seguintes afirmações são equivalentes:*

(i) A *é inversível.*

(ii) A *é linha-equivalente a* n × n *matriz unidade.*

(iii) A *é um produto de matrizes elementares.*

Demonstração. Seja R uma matriz linha-reduzida à forma em escada que seja linha-equivalente a A. Pelo Teorema 9 (ou seu corolário),

$$R = E_k \dots E_2 E_1 A$$

onde $E_1, \dots, E_k$ são matrizes elementares. Cada E_j é inversível e, portanto,

$$A = E_1^{-1} \dots E_k^{-1} R.$$

Como produtos de matrizes inversíveis são inversíveis, vemos que A é inversível se, e somente se, R é inversível. Como R é uma matriz (quadrada) linha-reduzida à forma em escada, R será inversível se, e somente se, cada linha de R contiver um elemento não nulo, isto é, se, e somente se, $R = I$. Mostramos assim que A é inversível se, e somente se, $R = I$, e, se $R = I$ então $A = E_k^{-1} \dots E_1^{-1}$. Deve ser evidente agora que (i), (ii) e (iii) são afirmações equivalentes sobre A.

Corolário. *Se* A *é uma* n × n *matriz inversível e se uma seqüência de operações elementares sobre linhas reduz* A *a unidade, então aquela mesma seqüência de operações sobre linhas quando aplicada a* I *produz* A^{-1}.

Corolário. *Sejam* A *e* B m × n *matrizes. Então* B *é linha-equivalente a* A *se, e somente se,* B = PA *onde* P *é uma* m × m *matriz inversível.*

Teorema 13. *Para uma* n × n *matriz* A, *as seguintes afirmações são equivalentes:*

(i) A *é inversível.*

(ii) *O sistema homogêneo* AX = 0 *possui somente a solução trivial* X = 0.

(iii) *O sistema de equações* AX = Y *possui uma solução* X *para toda* n × 1 *matriz* Y.

Demonstração. De acordo com o Teorema 7, a condição (ii) é equivalente ao fato de que A é linha-equivalente a matriz unidade. Pelo Teorema 12, (i) e (ii) são, portanto, equivalentes. Se A for

inversível, a solução de $AX = Y$ é $X = A^{-1}Y$. Reciprocamente, suponhamos que $AX = Y$ possua uma solução para cada Y. Seja R uma matriz linha-reduzida à forma em escada que seja linha-equivalente a A. Queremos mostrar que $R = I$. É suficiente mostrar que a última linha de R não é (identicamente) nula. Seja

$$E = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix}.$$

Se o sistema $RX = E$ puder ser resolvido em X, a última linha de R não pode ser nula. Sabemos que $R = PA$, onde P é inversível. Assim $RX = E$ se, e somente se, $AX = P^{-1}E$. De acordo com (iii) este último sistema tem uma solução.

Corolário. *Uma matriz quadrada com inversa à esquerda ou à direita é inversível.*

Demonstração. Seja A uma $n \times n$ matriz. Suponhamos que A possua uma inversa à esquerda, isto é, uma matriz B tal que $BA = I$. Então $AX = 0$ possui somente a solução trivial porque $X = IX = B(AX)$. Portanto, A é inversível. Por outro lado, suponhamos que A tenha uma inversa à direita, isto é, uma matriz C tal que $AC = I$. Então C possui uma inversa à esquerda e é, portanto, inversível. Segue daí que $A = C^{-1}$ e portanto A é inversível com inversa C.

Corolário. *Seja* $A = A_1A_2, \ldots, A_k$, *onde* $A_1, \ldots, A_k$ *são* $n \times n$ *matrizes (quadradas). Então* A *é inversível se, e somente se, cada* A_j *é inversível.*

Demonstração. Já demonstramos que o produto de duas matrizes inversíveis é inversível. A partir disto vê-se facilmente que se cada A_j é inversível então A é inversível.

Suponhamos agora que A seja inversível. Demonstremos primeiro que A_k é inversível. Suponhamos que X seja uma $n \times 1$ matriz e $A_kX = 0$. Então $AX = (A_1 \ldots A_{k-1})A_kX = 0$. Como A é inversível temos $X = 0$. Desta maneira, o sistema de equações $A_kX = 0$

não possui soluções não-triviais, portanto, A_k é inversível. Mas então $A_1 \ldots A_{k-1} = AA_k^{-1}$ é inversível. Pela razão anterior A_{k-1} é inversível. Prosseguindo desta forma, concluímos que cada A_j é inversível.

Gostaríamos de fazer um comentário final sobre a resolução de equações lineares. Suponhamos que A seja uma $m \times n$ matriz e que desejamos resolver o sistema de equações $AX = Y$. Se R é uma matriz linha-reduzida à forma em escada que é equivalente a A, então $R = PA$, onde P é uma $m \times n$ matriz inversível. As soluções do sistema $AX = Y$ são exatamente as soluções do sistema $RX = PY (= Z)$. Na prática, não é muito mais difícil determinar a matriz P do que linha-reduzir A a R. De fato, suponhamos que formemos a matriz completa A' do sistema $AX = Y$, com escalares arbitrários $y_1, \ldots, y_m$ na última coluna. Se agora efetuarmos sobre A' uma seqüência de operações elementares sobre linhas que reduza A a R, tornar-se-á evidente o que é a matriz P. (O leitor deve consultar o Exemplo 9 onde, em essência, aplicamos este processo.) Em particular, se A é uma matriz quadrada, este processo mostrará claramente se A é inversível ou não e, se A for inversível, qual é a inversa P. Como já apresentamos o núcleo de um exemplo deste tipo de cálculo, contentar-nos-emos com um exemplo 2×2.

Exemplo 15. Suponhamos que F seja o corpo dos números racionais e

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}.$$

Então

$$\begin{bmatrix} 2 & -1 & y_1 \\ 1 & 3 & y_2 \end{bmatrix} \xrightarrow{(3)} \begin{bmatrix} 1 & 3 & y_2 \\ 2 & -1 & y_1 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & 3 & y_2 \\ 0 & -7 & y_1 - 2y_2 \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 1 & 3 & y_2 \\ 0 & 1 & \frac{1}{7}(2y_2 - y_1) \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & 0 & \frac{1}{7}(y_2 + 3y_1) \\ 0 & 1 & \frac{1}{7}(2y_2 - y_1) \end{bmatrix}$$

onde se vê claramente que A é inversível e que

$$A^{-1} = \begin{bmatrix} \frac{3}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{2}{7} \end{bmatrix}.$$

Pode parecer muito trabalhoso continuar escrevendo os escalares arbitrários $y_1, y_2, \ldots,$ no cálculo da inversa. Algumas pessoas acham mais simples trabalhar com duas seqüências de matrizes, uma descrevendo a redução de A à matriz unidade e a outra, registrando o efeito da mesma seqüência de operações, começando com a matriz unidade. O leitor deverá julgar por si mesmo qual é o melhor processo.

Exemplo 16. Vamos determinar a inversa de

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & \frac{1}{12} & \frac{4}{45} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -\frac{1}{3} & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & 0 & \frac{1}{180} \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ \frac{1}{6} & -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -6 & 12 & 0 \\ 30 & -180 & 180 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} -9 & 60 & -60 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}.$$

Deve ter ocorrido ao leitor que fizemos uma longa discussão sobre linhas de matrizes e pouco dissemos sobre colunas. Concen-

tramos nossa atenção sobre as linhas porque isto pareceu mais natural do ponto de vista de equações lineares. Como não existe evidentemente nada sagrado sobre linhas, a discussão das últimas seções poderia muito bem ter sido feita usando-se colunas em vez de linhas. Se se define uma operação elementar sobre colunas e uma coluna equivalência de maneira análoga à operação elementar sobre linhas e à linha-equivalência é evidente que cada $m \times n$ matriz será coluna equivalente a uma matriz "coluna-reduzida à forma em escada". Além disso, cada operação elementar sobre colunas será da forma $A \to AE$, onde E é uma $n \times n$ matriz elementar e assim por diante.

Exercícios

1. Seja

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ -1 & 0 & 3 & 5 \\ 1 & -2 & 1 & 1 \end{bmatrix}.$$

Determinar uma matriz R linha-reduzida à forma em escada que seja linha-equivalente a A e uma 3×3 matriz inversível P tal que $R = PA$.

2. Fazer o Exercício 1, com

$$A = \begin{bmatrix} 2 & 0 & i \\ 1 & -3 & -i \\ i & 1 & 1 \end{bmatrix}.$$

3. Para cada uma das matrizes

$$\begin{bmatrix} 2 & 5 & -1 \\ 4 & -1 & 2 \\ 6 & 4 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & -1 & 2 \\ 3 & 2 & 4 \\ 0 & 1 & -2 \end{bmatrix}$$

usar operações elementares sobre linhas para descobrir se é inversível e, em caso afirmativo, determinar a inversa.

4. Seja

$$A = \begin{bmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{bmatrix}.$$

Para que X existe um escalar c tal que $AX = cX$?

5. Descobrir se

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

é inversível e determinar A^{-1} caso exista.

6. Suponhamos que A seja uma 2×1 matriz e que B seja uma 1×2 matriz. Demonstrar que $C = AB$ não é inversível.

7. Seja A uma $n \times n$ matriz (quadrada). Demonstrar as duas afirmações seguintes:

(a) Se A é inversível e $AB = 0$ para alguma $n \times n$ matriz B, então $B = 0$.

(b) Se A não é inversível, então existe uma $n \times n$ matriz B tal que $AB = 0$ mas $B \neq 0$.

8. Seja

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Demonstrar, usando operações elementares sobre linhas, que A é inversível se, e somente se, $(ad - bc) \neq 0$.

9. Uma $n \times n$ matriz A se diz triangular-superior se $A_{ij} = 0$ para $i > j$, isto é, se cada elemento abaixo da diagonal principal for 0. Demonstrar que uma matriz (quadrada) triangular superior é inversível se, e somente se, cada elemento da sua diagonal principal for diferente de 0.

10. Demonstrar a seguinte generalização do Exercício 6. Se A é uma $m \times n$ matriz, B é uma $n \times m$ matriz e $n < m$, então AB não é inversível.

11. Seja A uma $m \times n$ matriz. Mostrar que, por meio de um número finito de operações elementares sobre linhas e/ou colunas, pode-se passar de A a uma matriz R, "linha-reduzida à forma em escada" e "coluna-reduzida à forma em escada", isto é, $R_{ij} = 0$ se $i \neq j$, $R_{ii} = 1$, $1 \leq i \leq r$, $R_{ii} = 0$ se $i > r$. Mostrar que $R = PAQ$, onde P é uma $m \times m$ matriz inversível e Q é uma $n \times n$ matriz inversível.

12. O resultado do Exemplo 16 sugere que a matriz

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

é inversível e que A^{-1} possui elementos inteiros. Você saberia demonstrar este fato?

Capítulo 2
ESPAÇOS VETORIAIS

2.1 Espaços Vetoriais

Em várias partes da matemática, defrontamo-nos com um conjunto, tal que é, ao mesmo tempo, significativo e interessante lidar com "combinações lineares" dos objetos daquele conjunto. Por exemplo, em nosso estudo de equações lineares, foi bastante natural considerar combinações lineares das linhas de uma matriz. É provável que o leitor tenha estudado cálculo e tenha já lidado com combinações lineares de funções; isto certamente ocorreu se ele estudou equações diferenciais. Talvez o leitor tenha tido alguma experiência com vetores no espaço euclidiano tridimensional e, em particular, com combinações lineares de tais vetores.

A grosso modo, a álgebra linear é o ramo da matemática que trata das propriedades comuns a sistemas algébricos constituídos por um conjunto mais uma noção razoável de uma "combinação linear" de elementos do conjunto. Nesta seção definiremos o objeto matemático que, como a experiência mostrou, é a abstração mais útil deste tipo de sistema algébrico.

Definição. *Um* **espaço vetorial** (*ou espaço linear*) *consiste do seguinte*:

(1) *um corpo* F *de escalares*;

(2) *um corpo* V *de objetos, denominados vetores*;

(3) *uma regra* (*ou operação*), *dita adição de vetores, que associa a cada par de vetores* α, β *em* V *um vetor* $\alpha + \beta$ *em* V, *denominado a soma de* α *e* β, *de maneira tal que*:

(a) *a adição é comunicativa*, $\alpha + \beta = \beta + \alpha$;

(b) *a adição é associativa*, $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$;

(c) *existe um único vetor* 0 *em* V, *denominado o vetor nulo, tal que* $\alpha + 0 = \alpha$ para todo α em *V*;

(d) *para cada vetor* α *em* V *existe um único vetor* $-\alpha$ *em* V *tal que* $\alpha + (-\alpha) = 0$;

(4) *uma regra (ou operação), dita multiplicação escalar, que associa a cada escalar* c *em* F *e cada vetor* α *em* V *um vetor* $c\alpha$ *em* V, *denominado o produto de* c *por* α *de maneira tal que:*

(a) $1\alpha = \alpha$ *para todo* α *em* V;

(b) $(c_1c_2)\alpha = c_1(c_2\alpha)$;

(c) $c(\alpha + \beta) = c\alpha + c\beta$;

(d) $(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$.

É importante observar, como afirma a definição, que um espaço vetorial é um objeto composto de um corpo, um conjunto de "vetores" e duas operações com certas propriedades especiais. O mesmo conjunto de vetores pode ser parte de diversos espaços vetoriais (ver Exemplo 5 abaixo). Quando não há possibilidade de confusão, podemos simplesmente nos referir ao espaço vetorial por V ou, quando for desejável especificar o corpo, dizer que V é um **espaço vetorial sobre o corpo** F. O nome" vetor" é aplicado aos elementos do conjunto V mais por conveniência. A origem do nome é encontrada no Exemplo 1 abaixo, mas não se deve emprestar muita importância ao nome uma vez que a variedade de objetos que aparecem como sendo os vetores em V podem não apresentar muita semelhança com qualquer conceito de vetor adquirido *a priori* pelo leitor. Tentaremos indicar esta variedade através de uma lista de exemplos; nossa lista será consideravelmente ampliada assim que iniciarmos o estudo de espaços vetoriais.

Exemplo 1. O espaço das *n*-uplas, F^n. Seja F um corpo arbitrário e seja V o conjunto de todas as n-uplas $\alpha = (x_1, x_2, \ldots, x_n)$ de escalares x_i em F. Se $\beta = (y_1, y_2, \ldots, y_n)$ com y_i em F, a soma de α e β é definida por

$$\alpha + \beta = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n). \tag{2-1}$$

O produto de um escalar c por um vetor α é definido por

$$c\alpha = (cx_1, cx_2, \ldots, cx_n). \tag{2-2}$$

O fato de que esta adição de vetores e multiplicação escalar satisfazem as condições (3) e (4) é fácil de verificar, usando as propriedades semelhantes da adição e multiplicação de elementos de F.

Exemplo 2. O espaço das $m \times n$ matrizes, $F^{m \times n}$. Seja F um corpo arbitrário e sejam m e n inteiros positivos. Seja $F^{m \times n}$ o conjunto de todas as $m \times n$ matrizes sobre o corpo F. A soma de dois vetores A e B em $F^{m \times n}$ é definida por

(2-3) $$(A + B)_{ij} = A_{ij} + B_{ij}.$$

O produto de um escalar c pela matriz A é definido por

(2-4) $$(cA)_{ij} = cA_{ij}.$$

Observar que $F^{1 \times n} = F^n$.

Exemplo 3. O espaço das funções de um conjunto em um corpo. Seja F um corpo arbitrário e seja S um conjunto não-vazio arbitrário. Seja V o conjunto das funções do conjunto S em F. A soma de dois vetores f e g em V é o vetor $f + g$, isto é, a função de S em F, definida por

(2-5) $$(f + g)(s) = f(s) + g(s).$$

O produto do escalar c pela função f é a função cf definida por

(2-6) $$(cf)(s) = cf(s).$$

Os exemplos anteriores são os casos particulares deste. De fato, uma n-upla de elementos de F pode ser considerada como uma função do conjunto S dos inteiros $1, \ldots, n$ em F. Analogamente, uma $m \times n$ matriz sobre o corpo F é uma função do conjunto S de pares de inteiros (i, j), $1 \leq i \leq m$, $1 \leq j \leq n$, no corpo F. Para este terceiro exemplo indicaremos como se faz para verificar que as operações por nós definidas satisfazem as condições (3) e (4). Para a adição de vetores:

(a) Como a adição em F é comutativa,

$$f(s) + g(s) = g(s) + f(s)$$

para cada s em S, portanto as funções $f + g$ e $g + f$ são idênticas.

(b) Como a adição em F é associativa,

$$f(s) + [g(s) + h(s)] = [f(s) + g(s)] + h(s)$$

para cada s, portanto $f+(g+h)$ e $(f+g)+h$ são a mesma função.

(c) O único vetor nulo é a função nula que associa a cada elemento de S o escalar 0 em F.

(d) Para cada f em V, $(-f)$ é a função dada por

$$(-f)(s) = -f(s).$$

O leitor deverá achar fácil verificar que a multiplicação escalar satisfaz as condições de (4), fazendo como fizemos para a adição de vetores.

Exemplo 4. O espaço das funções polinomiais sobre um corpo F. Seja F um corpo e seja V o conjunto das funções f de F em F que são da forma

$$f(x) = c_0 + c_1 x + \ldots + c_n x^n \qquad (2\text{-}7)$$

onde $c_0, \ldots, c_n$ são escalares fixos em F (independentes de x). Uma função deste tipo é denominada uma **função polinomial sobre** F. Sejam a adição e multiplicação escalar definidas como no Exemplo 3. Deve-se observar aqui que se f e g são funções polinomiais e c está em F, então $f+g$ e cf são também funções polinomiais.

Exemplo 5. O corpo C dos números complexos pode ser considerado como um espaço vetorial sobre o corpo R dos números reais. De maneira mais geral, seja F o corpo dos números reais, e seja V o conjunto das n-uplas $\alpha = (x_1, \ldots, x_n)$ onde $x_1, \ldots, x_n$ são números *complexos*. Definamos a adição de vetores e a multiplicação escalar por (2-1) e (2-2), como no Exemplo 1. Desta forma obtemos um espaço vetorial sobre o corpo R que é bem diferente do espaço C^n e do espaço R^n.

Há alguns fatos simples que decorrem quase imediatamente da definição de um espaço vetorial e que passamos a deduzir. Se c é um escalar e 0 é o vetor nulo, então, por 3 (c) e 4 (c),

$$c0 = c(0 + 0) = c0 + c0.$$

Somando $-(c0)$ e usando 3(d) obtemos

$$c0 = 0. \qquad (2\text{-}8)$$

Analogamente, para o escalar 0 e qualquer vetor α temos que

$$0\alpha = 0. \tag{2-9}$$

Se c é um escalar não-nulo e α é um vetor tal que $c\alpha = 0$, então por (2-8), $c^{-1}(c\alpha) = 0$. Mas

$$c^{-1}(c\alpha) = (c^{-1}c)\alpha = 1\alpha = \alpha$$

logo, $\alpha = 0$. Assim, vemos que se c é um escalar e α um vetor tal que $c\alpha = 0$, então c é o escalar nulo ou α é o vetor nulo.

Se α é um vetor arbitrário em V, então

$$0 = 0\alpha = (1 - 1)\alpha = 1\alpha + (-1)\alpha = \alpha + (-1)\alpha$$

do que segue que

$$(-1)\alpha = -\alpha. \tag{2-10}$$

Finalmente, as propriedades associativa e comutativa da adição de vetores implicam que uma soma envolvendo um certo número de vetores é independente da maneira pela qual estes vetores são combinados ou associados. Por exemplo, se α_1, α_2, α_3, α_4 são vetores em V, então

$$(\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4) = [\alpha_2 + (\alpha_1 + \alpha_3)] + \alpha_4$$

e esta pode ser escrita sem confusão como

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4.$$

Definição. *Um vetor* β *em* V *é dito uma* **combinação linear** *dos vetores* $\alpha_1, \ldots, \alpha_n$ *em* V *se existem escalares* $c_1, \ldots, c_n$ *em* F *tais que*

$$\beta = c_1\alpha_1 + \ldots + c_n\alpha_n$$
$$= \sum_{i=1}^{n} c_i\alpha_i.$$

Outras extensões da propriedade associativa da adição de vetores e das propriedades distributivas 4 (c) e 4 (d) da multiplicação escalar aplicam-se a combinações lineares:

$$\sum_{i=1}^{n} c_i\alpha_i + \sum_{i=1}^{n} d_i\alpha_i = \sum_{i=1}^{n} (c_i + d_i)\alpha_i$$

$$c\sum_{i=1}^{n} c_i\alpha_i = \sum_{i=1}^{n} (cc_i)\alpha_i.$$

Certas partes da álgebra linear são intimamente relacionadas com a geometria. A própria palavra "espaço" sugere algo geométrico, como o faz a palavra "vetor" à maioria das pessoas. À medida que prossigamos nosso estudo de espaços vetoriais, o leitor observará que grande parte da terminologia possui uma conotação geométrica. Antes de concluirmos esta seção introdutória sobre espaços vetoriais, vamos considerar a relação dos espaços vetoriais com a geometria até um ponto que indique pelo menos a origem do nome "espaço vetorial". Esta será uma discussão breve e intuitiva.

Consideremos o espaço vetorial R^3. Na geometria analítica, identificamos as ternas (x_1, x_2, x_3) de números reais com os pontos do espaço euclidiano tridimensional. Naquele contexto, um vetor é usualmente definido como sendo um segmento de reta orientado PQ, que vai de um ponto P do espaço a outro ponto Q. Isto significa uma formulação cuidadosa da idéia da "flecha" de P a Q. Da forma como os vetores são usados, pretende-se que eles sejam determinados por seu comprimento, direção e sentido. Assim, é necessário identificar dois segmentos de reta orientados se eles têm o mesmo comprimento, direção e sentido.

O segmento de reta orientado PQ, que vai do ponto $P = (x_1, x_2, x_3)$ ao ponto $Q = (y_1, y_2, y_3)$, tem o mesmo comprimento, direção e sentido que o segmento de reta orientado que vai da origem $O = (0, 0, 0)$ ao ponto $(y_1 - x_1, y_2 - x_2, y_3 - x_3)$. Além disso, este é o único segmento que emana da origem e tem o mesmo comprimento, direção e sentido que PQ. Assim, se resolvermos estudar apenas os vetores que emanam da origem, existe exatamente um vetor associado a cada comprimento, direção e sentido dados.

O vetor OP, que vai da origem a $P = (x_1, x_2, x_3)$, é completamente determinado por P, portanto é possível identificar este vetor com o ponto P. Em nossa definição do espaço vetorial R^3, os vetores são definidos como sendo simplesmente as ternas (x_1, x_2, x_3).

Dados pontos $P = (x_1, x_2, x_3)$ e $Q = (y_1, y_2, y_3)$, a definição da soma dos vetores OP e OQ pode ser dada geometricamente. Se os vetores não são paralelos, então os segmentos OP e OQ determinam um plano e estes segmentos são dois dos lados de um paralelogramo naquele plano (ver Figura 1). Uma diagonal deste paralelogramo estende-se de O a um ponto S e a soma de

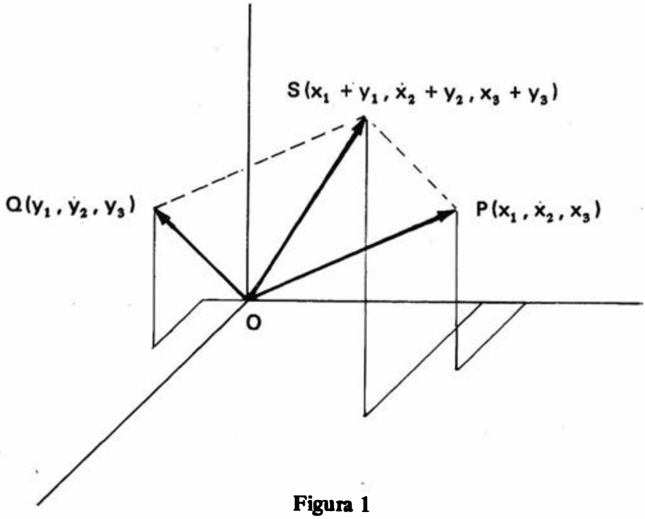

Figura 1

OP e OQ é definida como sendo o vetor OS. As coordenadas do ponto S são $(x_1 + y_1, x_2 + y_2, x_3 + y_3)$, logo esta definição geométrica da adição de vetores é equivalente à definição algébrica do Exemplo 1.

A multiplicação escalar tem uma interpretação geométrica mais simples. Se c é um número real, então, o produto de c pelo vetor OP é o vetor que parte da origem, tem comprimento $|c|$ vezes o comprimento de OP, mesma direção que OP e um sentido que concorda com o de OP se $c > 0$ e é oposto ao de OP se $c < 0$. Esta multiplicação escalar produz exatamente o vetor OT onde $T = (cx_1, cx_2, cx_3)$ e é portanto compatível com a definição algébrica dada para R^3.

De vez em quando, o leitor provavelmente achará útil "pensar geometricamente" sobre espaços vetoriais, isto é, desenhar figuras para uso próprio para ilustrar e motivar algumas idéias. Na verdade deve fazer isto. Contudo, ao fazer tais ilustrações, deve ter

em mente que, por estarmos tratando de espaços vetoriais como sistemas algébricos, todas as demonstrações que fizermos serão de natureza algébrica.

Exercícios

1. Se F é um corpo, verificar que F^n (tal como definido no Exemplo 1) é um espaço vetorial sobre o corpo F.
2. Se V é um espaço vetorial sobre o corpo F, verificar que

$$(\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4) = [\alpha_2 + (\alpha_3 + \alpha_1)] + \alpha_4$$

para todos vetores α_1, α_2, α_3 e α_4 em V.
3. Se C é o corpo dos números complexos, quais vetores em C^3 são combinações lineares de $(1, 0, -1)$, $(0, 1, 1)$ e $(1, 1, 1)$?
4. Seja V o conjunto de todos os pares (x, y) de números reais e seja F o corpo dos números reais. Definamos

$$(x, y) + (x_1, y_1) = (x + x_1, y + y_1)$$
$$c(x, y) = (cx, y).$$

V, com estas operações, é um espaço vetorial sobre o corpo dos números reais?
5. Em R^n definamos duas operações

$$\alpha \oplus \beta = \alpha - \beta$$
$$c \cdot \alpha = -c\alpha.$$

As operações do lado direito são as operações usuais. Quais dos axiomas de espaço vetorial são satisfeitos por $(R^n, \oplus, \cdot)$?

6. Seja V o conjunto de todas as funções f definidas sobre a reta real e tomando valores complexos, tais que (para todo t em R)

$$f(-t) = \overline{f(t)}.$$

A barra indica conjugação complexa. Mostrar que V, com as operações

$$(f + g)(t) = f(t) + g(t)$$
$$(cf)(t) = cf(t)$$

é um espaço vetorial sobre o corpo dos números *reais*. Dar um exemplo de uma função em V que não assuma somente valores reais.
7. Seja V o conjunto de todos os pares (x, y) de números reais e seja F o corpo dos números reais. Definamos

$$(x, y) + (x_1, y_1) = (3y + 3y_1, -x - x_1)$$
$$c(x, y) = (3cy, -cx).$$

Verificar se V, com estas operações, é um espaço vetorial.

2.2 Subespaços

Nesta seção introduziremos alguns conceitos básicos no estudo dos espaços vetoriais.

Definição. *Seja* V *um espaço vetorial sobre o corpo* F. *Um* **subespaço** *de* V *é um subconjunto* W *de* V *que é um espaço vetorial sobre* F *com as operações de adição de vetores e multiplicação escalar de* V.

Uma verificação direta dos axiomas para um espaço vetorial mostra que o subconjunto W de V é um subespaço se para todos α e β em W o vetor $\alpha+\beta$ está ainda em W; o vetor nulo está em W; para todo α em W o vetor $(-\alpha)$ está em W; para todo α em W e todo escalar c o vetor $c\alpha$ está em W. A comutatividade e associatividade da adição de vetores e as propriedades (4) (*a*), (*b*), (*c*) e (*d*) da multiplicação escalar não precisam ser verificadas, uma vez que são propriedades das operações em V. Podemos simplificar ainda mais as coisas.

Teorema 1. *Um subconjunto não-vazio* W *de* V *é um subespaço de* V *se, e somente se, para cada par de vetores* α, β *em* W *e cada escalar* c *em* F, *o vetor* $c\alpha+\beta$ *está em* W.

Demonstração. Suponhamos que W seja um subconjunto não-vazio de V tal que $c\alpha+\beta$ pertença a W para todos os vetores α, β em W e todos escalares c em F. Como W é não-vazio, existe um vetor ρ em W, logo $(-1)\rho+\rho=0$ está em W. Então se α é um vetor arbitrário em W e c um escalar arbitrário, o vetor $c\alpha=c\alpha+0$ está em W. Em particular $(-1)\alpha=-\alpha$ está em W. Finalmente se α e β estão em W, então $\alpha+\beta=1\alpha+\beta$ está em W. Assim, W é um subespaço de V.

Reciprocamente, se W é um subespaço de V, α e β estão em W e c é um escalar, certamente $c\alpha+\beta$ está em W.

Algumas pessoas preferem usar a propriedade $c\alpha+\beta$ do Teorema 1 como definição de subespaço. Isto faz pouca diferença. O ponto importante é que se W for um subconjunto não vazio de V tal que $c\alpha+\beta$ esteja em V para todos α, β em W e todos c em F, então W (com as operações herdadas de V) é um espaço vetorial. Este fato nos fornece muitos exemplos novos de espaços vetoriais.

Exemplo 6.

(a) Se V é um espaço vetorial arbitrário, V é um subespaço de V; o subconjunto constituído somente pelo vetor nulo é um

subespaço de V, denominado o **subespaço nulo** de V.

(b) Em F^n, o conjunto das n-uplas $(x_1, \ldots, x_n)$ com $x_1 = 0$ é um subespaço; contudo, o conjunto das n-uplas com $x_1 = 1 + x_2$ não é um subespaço $(n \geq 2)$.

(c) O espaço das funções polinomiais sobre o corpo F é um subespaço do espaço de todas as funções de F em F.

(d) Uma $n \times n$ matriz (quadrada) A sobre o corpo F é **simétrica** se $A_{ij} = A_{ji}$ para todos i e j. As matrizes simétricas formam um subespaço do espaço de todas as $n \times n$ matrizes sobre F.

(e) Uma $n \times n$ matriz (quadrada) A sobre o corpo C dos números complexos é **hermitiana** (ou **auto-adjunta**) se

$$A_{jk} = A_{kj}$$

para todos j, k, sendo que a barra indica conjugação complexa. Uma 2×2 matriz é hermitiana se, e somente se, é da forma

$$\begin{bmatrix} z & x + iy \\ x - iy & w \end{bmatrix}$$

onde x, y, z e w são números reais. O conjunto de todas as matrizes hermitianas *não* é um subespaço do espaço de todas as $n \times n$ matrizes sobre C. De fato, se A é hermitiana, todos os elementos A_{11}, $A_{22}, \ldots$, de sua diagonal são números reais mas os elementos diagonais de iA em geral não são reais. Por outro lado, verifica-se facilmente que o conjunto das $n \times n$ matrizes hermitianas complexas é um espaço vetorial sobre o corpo R dos números reais (com as operações usuais).

Exemplo 7. O espaço-solução de um sistema de equações lineares homogêneas. Seja A uma $m \times n$ matriz sobre F. Então o conjunto de todas as $n \times 1$ matrizes-(colunas) X sobre F tais que $AX = 0$ é um subespaço do espaço de todas as $n \times 1$ matrizes sobre F. Para demonstrar isto precisamos mostrar que $A(cX + Y) = 0$ para $AX = 0$, $AY = 0$ e c um escalar arbitrário em F. Isto decorre imediatamente do seguinte fato geral:

Lema. *Se* A *é uma* m × n *matriz sobre* F *e* B, C *são* n × p *matrizes sobre* F, *então*

$$A(dB + C) = d(AB) + AC \tag{2-11}$$

para todo escalar d *em* F.

Demonstração.
$$\begin{aligned}[A(dB + C)]_{ij} &= \sum_k A_{ik}(dB + C)_{kj} \\ &= \sum_k (dA_{ik}B_{kj} + A_{ik}C_{kj}) \\ &= d\sum_k A_{ik}B_{kj} + \sum_k A_{ik}C_{kj} \\ &= d(AB)_{ij} + (AC)_{ij} \\ &= [d(AB) + AC]_{ij}.\end{aligned}$$

Analogamente, pode-se mostrar que $(dB + C)A = d(BA) + CA$, se as somas e produtos de matrizes estão definidos.

Teorema 2. *Seja* V *um espaço vetorial sobre o corpo* F. *A interseção de uma coleção arbitrária de subespaços de* V *é um subespaço de* V.

Demonstração. Seja $\{W_a\}$ uma coleção de subespaços de V e seja $W = \bigcap_a W_a$ a sua interseção. Recordemos que W é definido como sendo o conjunto dos elementos pertencentes simultaneamente a W_a (ver Apêndice). Como cada W_a é um subespaço, todos contêm o vetor nulo. Assim o vetor nulo está na interseção W e W é não vazio. Sejam α e β vetores em W e seja c um escalar. Pela definição de W, tanto α como β pertencem a cada W_a e, como cada W_a é um subespaço, o vetor $(c\alpha + \beta)$ está em todo W_a. Assim, $(c\alpha + \beta)$ está em W. Pelo Teorema 1, W é um subespaço de V.

Do Teorema 2 decorre que se S é uma coleção arbitrária de vetores em V, então existe um menor subespaço de V que contém S, isto é, um subespaço que contém S e que está contido em todos os outros subespaços que contêm S.

Definição. *Seja* S *um conjunto de vetores num espaço vetorial* V. *O* **subespaço gerado** *por* S *é definido como sendo a interseção* W *de todos os subespaços de* V *que contêm* S. *Quando* S *é um conjunto finito de vetores,* $S = \{\alpha_1, \alpha_2, \ldots \alpha_n\}$, *denominaremos* W *simplesmente o* **subespaço gerado pelos vetores** $\alpha_1, \alpha_2, \ldots, \alpha_n$.

Teorema 3. *O subespaço gerado por um subconjunto não vazio* S *de um espaço vetorial* V *é o conjunto de todas as combinações lineares de vetores em* S.

Demonstração. Seja W o subespaço gerado por S. Então, cada combinação linear

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \ldots + x_m\alpha_m$$

de vetores $\alpha_1, \alpha_2, \ldots, \alpha_m$ em S evidentemente está em W. Assim, W contém o conjunto L de todas as combinações lineares de vetores em S. O conjunto L, por outro lado, contém S e é não vazio. Se α, β pertencem a L, então α é uma combinação linear,

$$\alpha = x_1\alpha_1 + x_2\alpha_2 + \ldots + x_m\alpha_m$$

de vetores α_i em S e β é uma combinação linear

$$\beta = y_1\beta_1 + y_2\beta_2 + \ldots + y_n\beta_n$$

de vetores β_j em S. Para cada escalar c,

$$c\alpha + \beta = \sum_{i=1}^{m} (cx_i)\alpha_i + \sum_{j=1}^{n} y_j\beta_j.$$

Logo $c\alpha + \beta$ pertence a L. Assim, L é um subespaço de V.

Mostramos acima que L é um subespaço de V que contém S e também que todo subespaço que contém S contém L. Decorre que L é a interseção de todos os subespaços que contêm S, isto é, que L é o subespaço gerado pelo conjunto S.

Definição. *Se* $\mathrm{S}_1, \mathrm{S}_2, \ldots, \mathrm{S}_k$ *são subconjuntos de um espaço vetorial* **V**, *o conjunto de todas as somas*

$$\alpha_1 + \alpha_2 + \ldots + \alpha_k$$

de vetores α_i *em* S_i *é dito a* **soma** *dos subconjuntos* $\mathrm{S}_1, \mathrm{S}_2, \ldots, \mathrm{S}_k$ *e é indicado por*

$$\mathrm{S}_1 + \mathrm{S}_2 + \ldots + \mathrm{S}_k$$

ou por

$$\sum_{i=1}^{k} \mathrm{S}_i.$$

Se $W_1, W_2, \ldots, W_k$ são subespaços de V, então vê-se facilmente que a soma

$$W = W_1 + W_2 + \dots + W_k$$

é um subespaço de V que contém cada um dos subespaços W_i. Disto decorre, como na demonstração do Teorema 3, que W é o subespaço gerado pela reunião de $W_1, W_2, \dots, W_k$.

Exemplo 8. Seja F um subcorpo do corpo C dos números complexos. Suponhamos que

$$\begin{aligned} \alpha_1 &= (1, 2, 0, 3, 0) \\ \alpha_2 &= (0, 0, 1, 4, 0) \\ \alpha_3 &= (0, 0, 0, 0, 1). \end{aligned}$$

Pelo Teorema 3, um vetor α está no subespaço W de F^5 gerado por $\alpha_1, \alpha_2, \alpha_3$ se, e somente se, existem escalares c_1, c_2, c_3 em F tais que

$$\alpha = c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3.$$

Portanto, W consiste de todos os vetores da forma

$$\alpha = (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3)$$

onde c_1, c_2, c_3 são escalares arbitrários em F. W pode ser descrito de outra forma como sendo o conjunto de todas as quíntuplas

$$\alpha = (x_1, x_2, x_3, x_4, x_5)$$

com x_i em F tais que

$$\begin{aligned} x_2 &= 2x_1 \\ x_4 &= 3x_1 + 4x_3. \end{aligned}$$

Assim, $(-3, -6, 1, -5, 2)$ está em W, enquanto que $(2, 4, 6, 7, 8)$ não está.

Exemplo 9. Seja F um subcorpo do corpo C dos números complexos e seja V o espaço vetorial das 2×2 matrizes sobre F. Seja W_1 o subconjunto de V constituído por todas as matrizes da forma

$$\begin{bmatrix} x & y \\ z & 0 \end{bmatrix}$$

onde x, y, z são escalares arbitrários em F. Finalmente, seja W_2 o subconjunto de V constituído por todas as matrizes da forma

$$\begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$$

onde x e y são escalares arbitrários em F. Então, W_1 e W_2 são subespaços de V. Além disso

$$V = W_1 + W_2$$

pois

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix}.$$

O subespaço $W_1 \cap W_2$ consiste de todas as matrizes da forma

$$\begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix}.$$

Exemplo 10. Seja A uma $m \times n$ matriz sobre um corpo F. Os **vetores-linhas** de A são os vetores em F^n dados por $\alpha_i = (A_{i1}, \ldots, A_{in})$, $i = 1, \ldots, m$. O subespaço de F^n gerado pelos vetores-linhas de A é denominado o **espaço-linha** de A. O subespaço considerado no Exemplo 8 é o espaço-linha da matriz.

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Ele também é o espaço-linha da matriz

$$B = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -4 & -8 & 1 & -8 & 0 \end{bmatrix}$$

Exemplo 11. Seja V o espaço das funções polinomiais sobre F. Seja S o subconjunto de V constituído pelas funções polinomiais $f_0, f_1, f_2, \ldots,$ definidas por

$$f_n(x) = x^n, \; n = 0, 1, 2, \dots$$

Então V é o subespaço gerado pelo conjunto S.

Exercícios

1. Quais dos seguintes conjuntos de vetores $\alpha = (a_1, \dots, a_n)$ em R^n são subespaços de R^n? ($n \geq 3$)

   (a) todos α tais que $a_1 \geq 0$;
   (b) todos α tais que $a_1 + 3a_2 = a_3$;
   (c) todos α tais que $a_2 = a_1^2$;
   (d) todos α tais que $a_1 a_2 = 0$;
   (e) todos α tais que a_2 seja racional.

2. Seja V o espaço vetorial (real) de todas as funções f de R em R. Quais dos seguintes conjuntos de funções são subespaços de V?

   (a) todas f tais que $f(x^2) = f(x)^2$;
   (b) todas f tais que $f(0) = f(1)$;
   (c) todas f tais que $f(3) = 1 + f(-5)$;
   (d) todas f tais que $f(-1) = 0$
   (e) todas f que são contínuas.

3. O vetor $(3.\; -1.\; 0.\; -1)$ está no subespaço de R^4 gerado pelos vetores $(2, -1, 3, 2)$, $(-1, 1, 1, -3)$ e $(1, 1, 9, -5)$?

4. Seja W o conjunto de todos os $(x_1, x_2, x_3, x_4, x_5)$ em R^5 que satisfazem

$$\begin{aligned} 2x_1 \quad x_2 + \tfrac{4}{3}x_3 - x_4 \qquad\quad &= 0 \\ x_1 \qquad + \tfrac{2}{3}x_3 \qquad\quad - x_5 &= 0 \\ 9x_1 - 3x_2 + 6x_3 - 3x_4 - 3x_5 &= 0. \end{aligned}$$

   Determinar um conjunto finito de vetores que gere W.

5. Seja F um corpo e seja n um inteiro positivo ($n \geq 2$). Seja V o espaço vetorial das $n \times n$ matrizes sobre F. Quais dos seguintes conjuntos de matrizes A em V são subespaços de V?

   (a) todas A inversíveis;
   (b) todas A não-inversíveis;
   (c) todas A tais que $AB = BA$, onde B é uma certa matriz fixa em V;
   (d) todas A tais que $A^2 = A$.

6. (a) Demonstrar que os únicos subespaços de R^1 são R^1 e o subespaço nulo.
   (b) Demonstrar que um subespaço de R^2 ou é R^2, ou é o subespaço nulo ou então consiste de todos os múltiplos escalares de um certo vetor fixo em R^2. (O último tipo de subespaço é (intuitivamente) uma reta pela origem.)
   (c) Você é capaz de descrever os subespaços de R^3?

7. Sejam W_1 e W_2 subespaços de um espaço vetorial V tais que a reunião de W_1 e W_2 também seja um subespaço. Demonstrar que um dos espaços W_i está contido no outro.

8. Seja V o espaço vetorial das funções de R em R; seja V_p o subconjunto das funções pares, $f(-x) = f(x)$; seja V_i o subconjunto das funções ímpares, $f(-x) = -f(x)$.

(a) Demonstrar que V_p e V_i são subespaços de V.
(b) Demonstrar que $V_p + V_i = V$.
(c) Demonstrar que $V_p \cap V_i = \{0\}$.

9. Sejam W_1 e W_2 subespaços de um espaço vetorial V tais que $W_1 + W_2 = V$ e $W_1 \cap W_2 = \{0\}$. Demonstrar que para cada vetor α em V existem vetores *bem determinados* α_1 em W_1 e α_2 em W_2, tais que $\alpha = \alpha_1 + \alpha_2$.

2.3 Bases e Dimensão

Passamos agora à tarefa de atribuir uma dimensão a certos espaços vetoriais. Apesar de associarmos usualmente "dimensão" a algo geométrico, precisamos encontrar uma definição algébrica adequada da dimensão de um espaço vetorial. Isto será feito através do conceito de uma base para o espaço.

Definição. *Seja* V *um espaço vetorial sobre* F. *Um subconjunto* S *de* V *é dito* **linearmente dependente** (*ou, simplesmente,* **dependente**) *se existem vetores distintos* $\alpha_1, \alpha_2, \ldots, \alpha_n$ *em* S *e escalares* $c_1, c_2, \ldots, c_n$ *em* F, *não todos nulos, tais que*

$$c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = 0.$$

Um conjunto que não é linearmente dependente é dito **linearmente independente.** *Se o conjunto* S *contém apenas um número finito de vetores* $\alpha_1, \alpha_2, \ldots, \alpha_n$ *dizemos, às vezes, que* $\alpha_1, \alpha_2, \ldots, \alpha_n$ *são dependentes* (*ou independentes*) *em vez de dizer que* S *é dependente* (*ou independente*).

Decorrem facilmente da definição as conseqüências seguintes:

1. Todo conjunto que contém um conjunto linearmente dependente é linearmente dependente.

2. Todo subconjunto de um conjunto linearmente independente é linearmente independente.

3. Todo conjunto que contém o vetor nulo é linearmente dependente, pois $1 \cdot 0 = 0$.

4. Um conjunto S de vetores é linearmente independente se, e somente se, todo subconjunto finito de S é linearmente indepen-

dente, isto é, se, e somente se, para quaisquer vetores distintos $\alpha_1, \ldots, \alpha_n$ em S, $c_1\alpha_1 + \ldots + c_n\alpha_n = 0$ implica que cada $c_i = 0$.

Definição. *Seja* V *um espaço vetorial. Uma* **base** *de* V *é um conjunto linearmente independente de vetores em* V *que gera o espaço* V. *O espaço vetorial* V *é de* **dimensão finita** *se ele possui uma base finita.*

Exemplo 12. Seja F um subcorpo do corpo dos números complexos. Em F^3 os vetores

$$\begin{aligned}
\alpha_1 &= (\ \ 3,\ 0,\ -3)\\
\alpha_2 &= (-1,\ 1,\ \ \ 2)\\
\alpha_3 &= (\ \ 4,\ 2,\ -2)\\
\alpha_4 &= (\ \ 2,\ 1,\ \ \ 1)
\end{aligned}$$

são linearmente dependentes, pois

$$2\alpha_1 + 2\alpha_2 - \alpha_3 + 0 \cdot \alpha_4 = 0.$$

Os vetores

$$\begin{aligned}
\varepsilon_1 &= (1,\ 0,\ 0)\\
\varepsilon_2 &= (0,\ 1,\ 0)\\
\varepsilon_3 &= (0,\ 0,\ 1)
\end{aligned}$$

são linearmente independentes.

Exemplo 13. Seja F um corpo e, em F^n, seja S o subconjunto constituído dos vetores $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ definidos por

$$\begin{aligned}
\varepsilon_1 &= (1,\ 0,\ 0,\ \ldots,\ 0)\\
\varepsilon_2 &= (0,\ 1,\ 0,\ \ldots,\ 0)\\
&\cdots\cdots\cdots\cdots\cdots\\
\varepsilon_n &= (0,\ 0,\ 0,\ \ldots,\ 1).
\end{aligned}$$

Sejam $x_1, x_2, \ldots, x_n$ escalares em F e coloquemos $\alpha = x_1\varepsilon_1 + x_2\varepsilon_2 + \ldots + x_n\varepsilon_n$. Então

$$\alpha = (x_1, x_2, \ldots, x_n). \tag{2-12}$$

Isto mostra que $\varepsilon_1, \ldots, \varepsilon_n$ geram F^n. Como $\alpha = 0$ se, e somente se, $x_1 = x_2 = \ldots = x_n = 0$, os vetores $\varepsilon_1, \ldots, \varepsilon_n$ são linearmente indepen-

dentes. O conjunto $S=\{\varepsilon_1, \ldots, \varepsilon_n\}$ é portanto uma base de F^n. Denominamos esta base particular a **base canônica** de F^n.

Exemplo 14. Seja P uma $n \times n$ matriz inversível com elementos no corpo F. Então as colunas $P_1, \ldots, P_n$ de P formam uma base do espaço das matrizes-coluna $F^{n\times 1}$. Vemos isto da seguinte maneira. Se X é uma matriz-coluna, então

$$PX = x_1P_1 + \ldots + x_nP_n.$$

Como $PX=0$ admite somente a solução trivial $X=0$, segue-se que $\{P_1, \ldots, P_n\}$ é um conjunto linearmente independente. Por que ele gera $F^{n\times 1}$? Seja Y uma matriz-coluna arbitrária. Se $X = P^{-1}Y$, então $Y = PX$, isto é,

$$Y = x_1P_1 + \ldots + x_nP_n.$$

Portanto, $\{P_1, \ldots, P_n\}$ é uma base de $F^{n\times 1}$.

Exemplo 15. Seja A uma $m \times n$ matriz e seja S o espaço solução do sistema homogêneo $AX=0$ (exemplo 7). Seja R uma matriz linha-reduzida à forma em escada que seja linha-equivalente a A. Então S é também o espaço solução do sistema $RX=0$. Se R tiver r linhas não nulas, então o sistema de equações $RX=0$ simplesmente exprime r das incógnitas $x_1, \ldots, x_n$ em função das (n-r) incógnitas restantes. Suponhamos que os primeiros elementos não nulos, das linhas não nulas, ocorram nas colunas $k_1, \ldots, k_r$. Seja J o conjunto dos n-r índices, distintos de $k_1, \ldots, k_r$:

$$J = \{1, \ldots, n\} - \{k_1, \ldots, k_r\}.$$

O sistema $RX=0$ tem a forma

$$\begin{array}{c} x_{k_1} + \sum_J c_{1j}x_j = 0 \\ \vdots \\ x_{k_r} + \sum_J c_{rj}x_j = 0 \end{array}$$

onde os c_{ij} são certos escalares. Todas as soluções são obtidas atribuindo valores (arbitrários) aos x_j com j em J e calculando os valores correspondentes de $x_{k_1}, \ldots, x_{k_r}$. Para cada j em J seja E_j

a solução obtida, colocando $x_j = 1$ e $x_i = 0$ para todos os demais i em J. Afirmamos que os $(n\text{-}r)$ vetores E_j, j em J, formam uma base do espaço solução.

Como a matriz coluna E_j possui um 1 na linha j e zeros nas linhas indexadas por outros elementos de J, o Exemplo 13 mostra que o conjunto destes vetores é linearmente independente. Este conjunto gera o espaço solução pela seguinte razão: se a matriz coluna T, com elementos $t_1, \ldots, t_n$ estiver no espaço solução, a matriz

$$N = \sum_J t_j E_j$$

também estará no espaço solução *e* será uma solução tal que $x_j = t_j$ para cada j em J. A solução com essa propriedade é única; portanto, $N = T$ e T está no espaço gerado pelos vetores E_j.

Exemplo 16. Daremos agora um exemplo de uma base infinita. Seja F um subcorpo do corpo dos números complexos e seja V o espaço das funções polinomiais sobre F. Lembremos que essas funções são as funções de F em F da forma

$$f(x) = c_0 + c_1 x + \ldots + c_n x^n.$$

Seja $f_k(x) = x^k$, $k = 0, 1, 2, \ldots$. O conjunto (infinito) $\{f_0, f_1, f_2, \ldots\}$ é uma base de V. É claro que o conjunto gera V porque a função f (acima) é

$$f = c_0 f_0 + c_1 f_1 + \ldots + c_n f_n.$$

O leitor deverá perceber que isto é virtualmente uma repetição da definição de função polinomial, isto é, uma função f de F em F é uma função polinomial se, e somente se, existirem um inteiro n e escalares $c_0, \ldots, c_n$ tais que $f = c_0 f_0 + \ldots + c_n f_n$. Por que as funções são independentes? Mostrar que o conjunto $\{f_0, f_1, f_2, \ldots\}$ é independente significa mostrar que cada um de seus subconjuntos finitos é independente. Será suficiente mostrar que, para cada n, o conjunto $\{f_0, \ldots, f_n\}$ é independente. Suponhamos que

$$c_0 f_0 + \ldots + c_n f_n = 0.$$

Isto significa que

$$c_0 + c_1 x + \ldots + c_n x^n = 0$$

para todo x em F; em outras palavras, todo x em F é uma raiz do polinômio $f(x) = c_0 + c_1x + \ldots + c_nx^n$. Assumiremos que o leitor saiba que um polinômio de grau n com coeficientes complexos não pode ter mais que n raízes distintas. Segue-se que $c_0 = c_1 = \ldots = = c_n = 0$.

Exibimos uma base infinita de V. Isto significa que V não é de dimensão finita? Embora a resposta seja afirmativa, ela não decorre imediatamente da definição, pois, pelo que sabemos até agora, V poderia também ter uma base finita. Essa possibilidade é facilmente eliminada. (Nós a eliminaremos, de um modo geral, no próximo teorema.) Consideremos um número finito de funções polinomiais $g_1, \ldots, g_r$. Haverá uma maior potência de x aparecendo (com coeficiente não nulo) em $g_1(x), \ldots, g_r(x)$. Se este expoente for k, obviamente $f_{k+1}(x) = x^{k+1}$ não estará no espaço gerado por $g_1, \ldots g_r$. Logo, V não é de dimensão finita.

Uma observação final sobre esse exemplo é oportuna. Bases infinitas nada têm a ver com "combinações lineares infinitas". O leitor que sentir um impulso irresistível de injetar séries de potências neste exemplo

$$\sum_{k=0}^{\infty} c_kx^k$$

deverá estudá-lo novamente, com cuidado. Se isto não curá-lo, ele deverá, de agora em diante, restringir sua atenção a espaços de dimensão finita.

Teorema 4. *Seja* V *um espaço vetorial gerado por um conjunto finito de vetores* $\beta_1, \beta_2, \ldots, \beta_m$. *Então, todo conjunto independente de vetores em* V *é finito e contém no máximo* m *elementos.*

Demonstração. Para demonstrar o teorema basta mostrar que todo subconjunto S de V que contém mais de m vetores é linearmente dependente. Seja S um tal conjunto. Em S existem vetores distintos $\alpha_1, \alpha_2, \ldots, \alpha_n$ com $n > m$. Como $\beta_1, \ldots, \beta_m$ geram V existem escalares A_{ij} em F tais que

$$\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i.$$

Para n escalares arbitrários $x_1, x_2, \ldots, x_n$ temos

$$x_1\alpha_1 + \ldots + x_n\alpha_n = \sum_{j=1}^{n} x_j\alpha_j$$

$$= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\beta_i$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{m} (A_{ij}x_j)\,\beta_i$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j \right) \beta_i.$$

Como $n > m$, o Teorema 6 do Capítulo 1 implica que existem escalares $x_1, x_2, \ldots, x_n$ não todos nulos, tais que

$$\sum_{j=1}^{n} A_{ij}x_j = 0, \quad 1 \leq i \leq m.$$

Logo, $x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n = 0$. Isto mostra que S é um conjunto linearmente dependente.

Corolário 1. *Se* V *é um espaço vetorial de dimensão finita, então duas quaisquer bases de* V *têm o mesmo número* (*finito*) *de elementos.*

Demonstração. Sendo de dimensão finita, V possui uma base finita

$$\{\beta_1, \beta_2, \ldots, \beta_m\}.$$

Pelo Teorema 4, toda base de V é finita e contém no máximo m elementos. Assim, se $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ é uma base, $n \leq m$. Pela mesma razão, $m \leq n$. Logo $m = n$.

Este corolário nos permite definir a **dimensão** de um espaço vetorial de dimensão finita como sendo o número de elementos numa base de V. Indicaremos a dimensão de um espaço vetorial V de dimensão finita por dim V. Isto nos permite reformular o Teorema 4 como segue.

Corolário 2. *Seja* V *um espaço vetorial de dimensão finita e seja* n = dim V. *Então:*

(a) *todo subconjunto de* V *que contém mais de* n *vetores é linearmente dependente;*

(b) *nenhum subconjunto de* V *contendo menos de* n *vetores pode gerar* V.

Exemplo 17. Se F é um corpo, a dimensão de F^n é n, pois a base canônica de F^n contém n vetores. O espaço das matrizes $F^{m \times n}$ tem dimensão mn. Isto deve estar claro por analogia com F^n, pois as mn matrizes que têm um 1 na posição i, j e zero nas demais, formam uma base de $F^{m \times n}$. Se A for uma $m \times n$ matriz, então o espaço solução para A tem dimensão $n - r$, onde r é o número de linhas não nulas de uma matriz linha-reduzida à forma em escada e linha-equivalente a A. Veja o Exemplo 15.

Se V for um espaço vetorial arbitrário sobre F, o subespaço nulo de V será gerado pelo vetor 0, mas $\{0\}$ é um conjunto linearmente dependente e não uma base. Por isso convencionaremos que o subespaço nulo tem dimensão 0. Poderíamos ter chegado à mesma conclusão, argumentando que o conjunto vazio é uma base do subespaço nulo. O conjunto vazio gera $\{0\}$ pois a interseção de todos os subespaços contendo o conjunto vazio é $\{0\}$ e o conjunto vazio é linearmente independente por não conter nenhum vetor.

Lema. *Seja* S *um subconjunto linearmente independente de um espaço vetorial* V. *Suponhamos que β seja um vetor em* V *que não esteja no subespaço gerado por* S. *Então o conjunto obtido acrescentando-se β a* S *é linearmente independente.*

Demonstração. Suponhamos que $\alpha_1, \ldots, \alpha_m$ sejam vetores distintos em S e que

$$c_1\alpha_1 + \ldots + c_m\alpha_m + b\beta = 0.$$

Então $b = 0$, caso contrário

$$\beta = \left(-\frac{c_1}{b}\right)\alpha_1 + \ldots + \left(-\frac{c_m}{b}\right)\alpha_m$$

e β estaria no subespaço gerado por S. Assim, $c_1\alpha_1 + \ldots + c_m\alpha_m = 0$, e como S é um conjunto linearmente independente, cada $c_i = 0$.

Teorema 5. *Se* W *é um subespaço de um espaço vetorial* V *de dimensão finita, todo subconjunto de* W *que é linearmente independente é finito e é parte de uma base (finita) de* W.

Demonstração. Suponhamos que S_0 seja um subconjunto de W linearmente independente. Se S é um subconjunto de W linearmente independente contendo S_0, então S também é um subcon-

junto de W linearmente independente; como V é de dimensão finita, S contém no máximo dim V elementos.

Estendemos S_0 até obtermos uma base de W, da seguinte maneira. Se S_0 gerar W, então S_0 é uma base de W e a demonstração está terminada. Se S_0 não gerar W usamos o lema anterior para encontrar um vetor β_1 em W tal que o conjunto $S_1 = S_0 \cup \{\beta_1\}$ seja independente. Se S_1 gerar W, ótimo. Caso contrário, aplicamos o lema para obter um vetor β_2 em W tal que $S_2 = S_1 \cup \{\beta_2\}$ seja independente. Continuando dessa maneira (em no máximo dim V passagens) obtemos o conjunto

$$S_m = S_0 \cup \{\beta_1, \ldots, \beta_m\}$$

que é uma base de W.

Corolário 1. *Se* W *é um subespaço próprio de um espaço vetorial* V *de dimensão finita, então* W *é de dimensão finita e dim* W < < *dim* V .

Demonstração. Podemos supor que W contém um vetor $\alpha \neq 0$. Pelo Teorema 5 e sua demonstração, existe uma base de W que contém α e no máximo dim V elementos. Logo W é de dimensão finita e dim $W \leq$ dim V. Como W é subespaço próprio, existe um vetor β em V que não está em W. Acrescentando β a uma base arbitrária de W obtemos um subconjunto de V linearmente independente. Portanto dim $W <$ dim V.

Corolário 2. *Num espaço vetorial* V *de dimensão finita todo conjunto não-vazio de vetores linearmente independentes é parte de uma base.*

Corolário 3. *Seja* A *uma* n × n *matriz sobre um corpo* F *e suponhamos que os vetores-linhas de* A *formem um conjunto de valores de* F^n *linearmente independentes. Então* A *é inversível.*

Demonstração. Sejam $\alpha_1, \alpha_2, \ldots, \alpha_n$ os vetores-linhas de A e suponhamos que W seja o subespaço de F^n gerado por $\alpha_1, \alpha_2, \ldots, \alpha_n$. Como $\alpha_1, \alpha_2, \ldots, \alpha_n$ são linearmente independentes, a dimensão de W é n. O Corolário 1 mostra agora que $W = F^n$. Logo, existem escalares B_{ij} em F tais que

$$\varepsilon_i = \sum_{j=1}^{n} B_{ij}\alpha_j, \qquad 1 \leq i \leq n$$

onde $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ é a base canônica de F^n. Portanto, para a matriz B com elementos B_{ij}, temos

$$BA = I.$$

Teorema 6. *Se* W$_1$ *e* W$_2$ *são subespaços de dimensão finita de um espaço vetorial* V, *então* W$_1$ + W$_2$ *é de dimensão finita e*

$$\dim W_1 + \dim W_2 = \dim (W_1 \cap W_2) + \dim (W_1 + W_2).$$

Demonstração. Pelo Teorema 5 e seus corolários, $W_1 \cap W_2$ tem uma base finita $\{\alpha_1, \ldots, \alpha_k\}$ que é parte de uma base

$$\{\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m\} \text{ de } W_1$$

e parte de uma base

$$\{\alpha_1, \ldots, \alpha_k, \gamma_1, \ldots, \gamma_n\} \text{ de } W_2.$$

O subespaço $W_1 + W_2$ é gerado pelos vetores

$$\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m, \gamma_1, \ldots, \gamma_n$$

e estes vetores formam um conjunto independente. De fato, suponhamos que

$$\Sigma\, x_i\alpha_i + \Sigma\, y_j\beta_j + \Sigma\, z_r\gamma_r = 0.$$

Então

$$-\Sigma\, z_r\gamma_r = \Sigma\, x_i\alpha_i + \Sigma\, y_j\beta_j$$

o que mostra que $\Sigma\, z_r\gamma_r$ pertence a W_1. Como $\Sigma\, z_r\gamma_r$ pertence também a W_2, segue que

$$\Sigma\, z_r\gamma_r = \Sigma\, c_i\alpha_i$$

para certos escalares $c_1, \ldots, c_k$. Por ser o conjunto

$$\{\alpha_1, \ldots, \alpha_k, \gamma_1, \ldots, \gamma_n\}$$

independente, cada um dos escalares $z_r = 0$. Portanto,

$$\Sigma x_i\alpha_i + \Sigma y_j\beta_j = 0$$

e como

$$\{\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m\}$$

também é um conjunto independente, cada $x_i = 0$ e cada $y_j = 0$. Assim,

$$\{\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m, \gamma_1, \ldots, \gamma_n\}$$

é uma base de $W_1 + W_2$. Finalmente

$$\begin{aligned} \dim W_1 + \dim W_2 &= (k + m) + (k + n) \\ &= k + (m + k + n) \\ &= \dim (W_1 \cap W_2) + \dim (W_1 + W_2). \end{aligned}$$

Encerremos esta seção com uma observação a respeito de dependência e independência linear. Esses conceitos foram definidos para conjuntos de vetores. É útil defini-los para seqüências finitas (n-uplas ordenadas) de vetores: $\alpha_1, \ldots, \alpha_n$. Dizemos que os vetores $\alpha_1, \ldots, \alpha_n$ são **linearmente dependentes** se existirem escalares $c_1, \ldots, c_n$, não todos nulos, tais que $c_1\alpha_1 + \ldots + c_n\alpha_n = 0$. Isso tudo é tão natural que o leitor talvez descubra já ter usado esta terminologia. Qual é a diferença entre uma seqüência finita $\alpha_1, \ldots, \alpha_n$ e um conjunto $\{\alpha_1, \ldots, \alpha_n\}$? Existem duas diferenças, identidade e ordem.

Ao discutirmos o conjunto $\{\alpha_1, \ldots, \alpha_n\}$ normalmente admitimos que os vetores $\alpha_1, \ldots, \alpha_n$ sejam distintos dois a dois. Em uma seqüência $\alpha_1, \ldots, \alpha_n$ todos os α_i podem ser o mesmo vetor. Se $\alpha_i = \alpha_j$ para algum $i \neq j$, então a seqüência $\alpha_1, \ldots, \alpha_n$ é linearmente dependente:

$$\alpha_i + (-1)\alpha_j = 0.$$

Assim, se $\alpha_1, \ldots, \alpha_n$ forem linearmente independentes, eles são distintos e podemos falar a respeito do conjunto $\{\alpha_1, \ldots, \alpha_n\}$ e saber que ele possui n vetores. Logo é claro que não surgirá nenhuma confusão ao discutirmos bases e dimensão. A dimensão de um espaço V de dimensão finita é o maior n tal que alguma n-upla de vetores em V seja linearmente independente — e assim por diante.

O leitor que achar que esse parágrafo faz muito alarde a respeito de nada, poderá perguntar a si mesmo se os vetores

$$\alpha_1 = (e^{\pi/2},\ 1)$$
$$\alpha_2 = (\sqrt[3]{110},\ 1)$$

são linearmente independentes em R^2.

Os elementos de uma seqüência são enumerados em uma ordem específica. Um conjunto é uma coleção de objetos sem nenhum arranjo especial ou ordem. Naturalmente, para descrevermos o conjunto podemos enumerar seus elementos e isso requer a escolha de uma ordem. Mas, a ordem não faz parte do conjunto. Os conjuntos $\{1, 2, 3, 4\}$ e $\{4, 3, 2, 1\}$ são idênticos, enquanto que a seqüência $1, 2, 3, 4$ é bem diferente da seqüência $4, 3, 2, 1$. O aspecto de ordem da seqüência não tem efeito sobre questões de independência, dependência etc., porque dependência (como definida) não é afetada pela ordem. A seqüência $\alpha_n, \ldots, \alpha_1$ é dependente se, e somente se, a seqüência $\alpha_1, \ldots, \alpha_n$ é dependente. Na próxima seção, a ordem será importante.

Exercícios

1. Demonstrar que, se dois vetores são linearmente dependentes, um deles é um múltiplo escalar do outro.
2. Os vetores

$$\alpha_1 = (1,\ 1,\ 2,\ 4) \qquad \alpha_2 = (2,\ -1,\ -5,\ 2)$$
$$\alpha_3 = (1,\ -1,\ -4,\ 0) \qquad \alpha_4 = (2,\ 1,\ 1,\ 6)$$

são linearmente independentes em R^4?

3. Determinar uma base do subespaço de R^4 gerado pelos quatro vetores do Exercício 2.

4. Mostrar que os vetores

$$\alpha_1 = (1,\ 0,\ -1), \qquad \alpha_2 = (1,\ 2,\ 1), \qquad \alpha_3 = (0,\ -3,\ 2)$$

formam uma base de R^3. Exprimir cada um dos vetores da base canônica como combinações lineares de α_1, α_2, α_3.

5. Determinar três vetores em R^3 que sejam linearmente dependentes e tais que dois quaisquer deles sejam linearmente independentes.

6. Seja V o espaço vetorial das 2×2 matrizes sobre o corpo F. Demonstrar que V tem dimensão 4 exibindo uma base de V que tenha 4 elementos.

7. Seja V o espaço vetorial do Exercício 6. Seja W_1 o conjunto das matrizes da forma

$$\begin{bmatrix} x & -x \\ y & z \end{bmatrix}$$

e seja W_2 o conjunto das matrizes da forma

$$\begin{bmatrix} a & b \\ -a & c \end{bmatrix}.$$

(a) Demonstrar que W_1 e W_2 são subespaços de V.

(b) Determinar as dimensões de W_1, W_2, $W_1 + W_2$ e $W_1 \cap W_2$.

8. Novamente, seja V o espaço das 2×2 matrizes sobre F. Determinar uma base $\{A_1, A_2, A_3, A_4\}$ de V tal que $A_j^2 = A_j$ para cada j.

9. Seja V um espaço vetorial sobre um subcorpo F do corpo dos números complexos. Suponhamos que α, β e γ sejam vetores de V linearmente independentes. Demonstrar que $(\alpha + \beta)$, $(\beta + \gamma)$ e $(\gamma + \alpha)$ são linearmente independentes.

10. Seja V um espaço vetorial sobre o corpo F. Suponhamos que exista um número finito de vetores $\alpha_1, \ldots, \alpha_r$ de V que gerem V. Demonstrar que V é de dimensão finita.

11. Seja V o conjunto das 2×2 matrizes A com elementos *complexos* satisfazendo $A_{11} + A_{22} = 0$.

(a) Mostrar que V é um espaço vetorial sobre o corpo dos números *reais*, com as operações usuais de adição de matrizes e multiplicação de uma matriz por um escalar.

(b) Determinar uma base desse espaço vetorial.

(c) Seja W o conjunto de todas as matrizes A em V tais que $A_{21} = -\bar{A}_{12}$ (a barra indica conjugação complexa). Demonstrar que W é um subespaço de V e determinar uma base de W.

12. Demonstrar que o espaço das $m \times n$ matrizes sobre o corpo F tem dimensão mn, exibindo uma base para este espaço.

13. Discutir o Exercício 9, para o caso em que V é um espaço vetorial sobre o corpo formado por dois elementos descritos no Exercício 5, Seção 1.1.

14. Seja V o conjunto dos números reais. Consideremos V como um espaço vetorial sobre o corpo dos números *racionais*, com as operações usuais. Demonstrar que este espaço vetorial *não é* de dimensão finita.

2.4 Coordenadas

Uma das caracteristicas úteis de uma base $\mathscr{B}$ de um espaço n-dimensional V é essencialmente que ela nos permite introduzir coordenadas em V análogas às "coordenadas naturais" x_i de um vetor $\alpha = (x_1, \ldots, x_n)$ do espaço F^n. Deste modo, as coordenadas de um vetor α de V em relação à base $\mathscr{B}$ serão os escalares que

servem para exprimir α como uma combinação linear dos vetores da base. Assim, gostaríamos de considerar as coordenadas naturais de um vetor α de F^n como sendo definidas por α e pela base canônica de F^n; contudo, ao adotarmos este ponto de vista precisamos ter um certo cuidado. Se

$$\alpha = (x_1, \ldots, x_n) = \Sigma\, x_i \varepsilon_i$$

e $\mathscr{B}$ é a base canônica de F^n, como são as coordenadas de α determinadas por $\mathscr{B}$ e α? Uma maneira de formular a resposta é esta: Um dado vetor α é expresso de maneira única como uma combinação linear dos vetores da base canônica, e a i-ésima coordenada x_i de α é o coeficiente de ε_i nesta expressão. Sob este ponto de vista podemos dizer qual é a i-ésima coordenada, pois temos uma ordenação "natural" dos vetores da base canônica, isto é, temos uma regra para determinar qual é o "primeiro" vetor da base, qual é o "segundo" e assim por diante. Se $\mathscr{B}$ é uma base arbitrária do espaço n-dimensional V, não teremos provavelmente nenhuma ordenação natural para os vetores de $\mathscr{B}$ e será portanto necessário impormos uma certa ordem sobre esses vetores antes de podermos definir "a i-ésima coordenada de α em relação a $\mathscr{B}$". Em outras palavras, coordenadas são definidas relativamente a seqüências de vetores e não em relação a conjuntos de vetores.

Definição. *Se* V *é um espaço vetorial de dimensão finita, uma* **base ordenada** *de* V *é uma seqüência finita de vetores linearmente independentes, que gera* V.

Se a seqüência $\alpha_1, \ldots, \alpha_n$ é uma base ordenada de V, então o conjunto $\{\alpha_1, \ldots, \alpha_n\}$ é uma base de V. A base ordenada é o conjunto, juntamente com uma ordem especificada. Vamos nos permitir um ligeiro abuso de notação e descrever tudo isto dizendo que

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$$

é uma base ordenada de V.

Suponhamos agora que V seja um espaço vetorial de dimensão finita sobre o corpo F e que

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$$

seja uma base ordenada de V. Dado α em V, existe uma única n-upla $(x_1, \ldots, x_n)$ de escalares tal que

$$\alpha = \sum_{i=1}^{n} x_i\alpha_i.$$

A n-upla é única, pois, se tivéssemos

$$\alpha = \sum_{i=1}^{n} z_i\alpha_i$$

então

$$\sum_{i=1}^{n} (x_i - z_i)\alpha_i = 0$$

e a independência linear dos α_i nos diria que $x_i - z_i = 0$ para cada i. Denominaremos x_i a i-ésima **coordenada de α em relação à base ordenada**

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}.$$

Se

$$\beta = \sum_{i=1}^{n} y_i\alpha_i$$

então

$$\alpha + \beta = \sum_{i=1}^{n} (x_i + y_i)\alpha_i$$

de modo que a i-ésima coordenada de $(\alpha + \beta)$ em relação a esta base ordenada é $(x_i + y_i)$. Analogamente, a i-ésima coordenada de $(c\alpha)$ é cx_i. Devemos também notar que toda n-upla $(x_1, \ldots, x_n)$ de F^n é a n-upla de coordenadas de algum vetor de V, a saber, o vetor

$$\sum_{i=1} x_i\alpha_i.$$

Resumindo, cada base ordenada de V determina uma correspondência bijetora

$$\alpha \rightarrow (x_1, \ldots, x_n)$$

entre o conjunto dos vetores de V e o conjunto das n-uplas de F^n. Esta correspondência tem a propriedade de que o correspondente de $(\alpha + \beta)$ é a soma em F^n dos correspondentes de α e β, e que o correspondente de $(c\alpha)$ é o produto em F^n do escalar c pelo correspondente de α.

Poder-se-ia perguntar neste ponto por que não tomar simplesmente uma base ordenada de V e descrever cada vetor de V por sua correspondente n-upla de coordenadas, visto que teríamos então a conveniência de operar apenas com n-uplas. Isto faria malograr nosso objetivo, por duas razões. Primeiro, como indica a nossa definição axiomática de espaço vetorial, estamos tentando aprender a raciocinar com espaços vetoriais como sistemas algébricos abstratos. Segundo, mesmo nos casos em que usamos coordenadas, os resultados importantes decorrem de nossa habilidade de mudar o sistema de coordenadas, isto é, mudar a base ordenada.

Freqüentemente o mais conveniente será usar a **matriz das coordenadas de α em relação à base ordenada $\mathscr{B}$**:

$$X = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

em vez da n-upla $(x_1, \ldots, x_n)$ das coordenadas. Para indicar que esta matriz de coordenadas depende da base, usaremos o símbolo

$$[\alpha]_{\mathscr{B}}$$

para a matriz das coordenadas do vetor α em relação à base ordenada $\mathscr{B}$. Esta notação será particularmente útil ao passarmos agora a descrever o que acontece com as coordenadas de um vetor α quando passamos de uma base ordenada a outra.

Suponhamos então que V seja n-dimensional e que

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{e} \quad \mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

sejam duas bases ordenadas de V. Existem escalares P_{ij}, bem determinados, tais que

$$\text{(2-13)} \qquad \alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i, \qquad 1 \leq j \leq n.$$

Sejam $x'_1, \ldots, x'_n$ as coordenadas de um dado vetor α em relação à base ordenada $\mathscr{B}'$. Então

$$\begin{aligned}\alpha &= x'_1\alpha'_1 + \ldots + x'_n\alpha'_n \\ &= \sum_{j=1}^{n} x'_j\alpha'_j \\ &= \sum_{j=1}^{n} x'_j \sum_{i=1}^{n} P_{ij}\alpha_i \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} (P_{ij}x'_j)\,\alpha_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij}x'_j\right) \alpha_i.\end{aligned}$$

Portanto, obtemos a relação

$$\alpha = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij}x'_j\right)\alpha_i. \tag{2-14}$$

Como as coordenadas $x_1, x_2, \ldots, x_n$ de α em relação à base ordenada $\mathscr{B}$ são determinadas de modo único, decorre de (2-14) que

$$x_i = \sum_{j=1}^{n} P_{ij}x'_j, \qquad 1 \leq i \leq n. \tag{2-15}$$

Seja P a $n \times n$ matriz cujo elemento i, j é o escalar P_{ij} e sejam X e X' as matrizes das coordenadas do vetor α em relação às bases ordenadas $\mathscr{B}$ e $\mathscr{B}'$. Podemos então reformular (2-15) como

$$X = PX'. \tag{2-16}$$

Como $\mathscr{B}$ e $\mathscr{B}'$ são conjuntos linearmente independentes, $X = 0$ se, e somente se, $X' = 0$. Assim, de (2-16) e do Teorema 7 do Capítulo 1, decorre que P é inversível. Logo,

$$X' = P^{-1}X. \tag{2-17}$$

Se usarmos a notação acima introduzida para a matriz das coordenadas de um vetor em relação a uma base ordenada, então (2-16) e (2-17) afirmam que

$$[\alpha]_{\mathscr{B}} = P[\alpha]_{\mathscr{B}'}$$

e

$$[\alpha]_{\mathscr{B}'} = P^{-1}[\alpha]_{\mathscr{B}} .$$

Portanto, a discussão precedente pode ser resumida como segue.

Teorema 7. *Seja* V *um espaço vetorial* n-*dimensional sobre o corpo* F *e sejam* $\mathscr{B}$ *e* $\mathscr{B}'$ *duas bases ordenadas de* V. *Então existe uma única* n × n *matriz* P, *necessariamente inversível, com elementos em* F, *tal que*

(i) $$[\alpha]_{\mathscr{B}} = P[\alpha]_{\mathscr{B}'}$$
(ii) $$[\alpha]_{\mathscr{B}'} = P^{-1}[\alpha]_{\mathscr{B}}$$

para todo vetor α *em* V. *As colunas de* P *são dadas por*

$$P_j = [\alpha'_j]_{\mathscr{B}} \qquad j = 1, \ldots, n.$$

Para completar a análise acima demonstraremos também o resultado que segue.

Teorema 8. *Suponhamos que* P *seja uma* n × n *matriz inversível sobre* F. *Seja* V *um espaço vetorial* n-*dimensional sobre* F *e seja* $\mathscr{B}$ *uma base ordenada de* V. *Então, existe uma única base ordenada* $\mathscr{B}'$ *de* V *tal que*

(i) $$[\alpha]_{\mathscr{B}} = P[\alpha]_{\mathscr{B}'}$$
(ii) $$[\alpha]_{\mathscr{B}'} = P^{-1}[\alpha]_{\mathscr{B}}$$

para todo vetor α *em* V.

Demonstração. Seja $\mathscr{B}$ constituída pelos vetores $\alpha_1, \ldots, \alpha_n$. Se $\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ é uma base ordenada de V para a qual (i) é válida, é claro que

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i .$$

Assim, basta-nos mostrar que os vetores α'_j, definidos por estas equações, formam uma base. Seja $Q = P^{-1}$. Então

$$\begin{aligned}\sum_j Q_{jk}\alpha'_j &= \sum_j Q_{jk} \sum_i P_{ij}\alpha_i \\ &= \sum_j \sum_i P_{ij}Q_{jk}\, \alpha_i \\ &= \sum_i \left(\sum_j P_{ij}Q_{jk}\right)\alpha_i \\ &= \alpha_k.\end{aligned}$$

Portanto, o subespaço gerado pelo conjunto

$$\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

contém $\mathscr{B}$, logo é igual a V. Assim, $\mathscr{B}'$ é uma base e, de sua definição e do Teorema 7, é evidente que (i) é válida, logo (ii) também o é.

Exemplo 18. Seja F um corpo e seja

$$\alpha = (x_1, x_2, \ldots, x_n)$$

um vetor de F^n. Se $\mathscr{B}$ é a base ordenada canônica de F^n,

$$\mathscr{B} = \{\varepsilon_1, \ldots, \varepsilon_n\},$$

a matriz das coordenadas do vetor α em relação à base $\mathscr{B}$ é dada por

$$[\alpha]_{\mathscr{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}.$$

Exemplo 19. Seja R o corpo dos números reais e seja θ um número real fixo. A matriz

$$P = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$$

é inversível e sua inversa é

$$P^{-1} = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

Portanto, para cada θ, o conjunto $\mathscr{B}'$ constituído pelos vetores $(\cos\theta, \operatorname{sen}\theta)$, $(-\operatorname{sen}\theta, \cos\theta)$ é uma base de R^2. intuitivamente esta base pode ser descrita como sendo a base obtida pela rotação de um ângulo θ da base canônica. Se α é o vetor (x_1, x_2), então

$$[\alpha]_{\mathscr{B}} = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

ou

$$\begin{aligned} x_1' &= x_1 \cos\theta + x_2 \operatorname{sen}\theta \\ x_2' &= -x_1 \operatorname{sen}\theta + x_2 \cos\theta. \end{aligned}$$

Exemplo 20. Seja F um subcorpo do corpo dos números complexos. A matriz

$$P = \begin{bmatrix} -1 & 4 & 5 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{bmatrix}$$

é inversível e sua inversa é

$$P^{-1} = \begin{bmatrix} -1 & 2 & \frac{11}{8} \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{bmatrix}.$$

Portanto, os vetores

$$\begin{aligned} \alpha_1' &= (-1, \ 0, \ 0) \\ \alpha_2' &= (\ 4, \ 2, \ 0) \\ \alpha_3' &= (\ 5, \ -3, \ 8) \end{aligned}$$

formam uma base $\mathscr{B}'$ de F^3. As coordenadas x_1', x_2', x_3' do vetor $\alpha = (x_1, x_2, x_3)$ em relação à base $\mathscr{B}'$ são dadas por

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} -x_1 + 2x_2 + \frac{11}{8} x_3 \\ \frac{1}{2} x_2 + \frac{3}{16} x_3 \\ \frac{1}{8} x_3 \end{bmatrix} = \begin{bmatrix} -1 & 2 & \frac{11}{8} \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Em particular,

$$(3, 2, -8) = -10\alpha'_1 - \tfrac{1}{2}\alpha'_2 - \alpha'_3.$$

Exercícios

1. Mostrar que os vetores

$$\alpha_1 = (1, 1, 0, 0), \qquad \alpha_2 = (0, 0, 1, 1)$$
$$\alpha_3 = (1, 0, 0, 4), \qquad \alpha_4 = (0, 0, 0, 2)$$

formam uma base de R^4. Determinar as coordenadas de cada um dos vetores da base canônica em relação à base ordenada $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$.

2. Determinar a matriz das coordenadas do vetor (1, 0, 1) em relação à base de C^3 constituída pelos vetores $(2i, 1, 0)$, $(2, -1, 1)$, $(0, 1 + i, -i)$, nesta ordem.

3. Seja $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ a base ordenada de R^3 constituída por

$$\alpha_1 = (1, 0, -1), \qquad \alpha_2 = (1, 1, 1), \qquad \alpha_3 = (1, 0, 0).$$

Quais são as coordenadas do vetor (a, b, c) em relação à base ordenada $\mathscr{B}$?

4. Seja W o subespaço de C^3 gerado por $\alpha_1 = (1, 0, i)$ e $\alpha_2 = (1 + i, 1, -1)$.

(a) Mostrar que α_1 e α_2 formam uma base de W.

(b) Mostrar que os vetores $\beta_1 = (1, 1, 0)$ e $\beta_2 = (1, i, 1 + i)$ estão em W e formam outra base de W.

(c) Quais são as coordenadas de α_1 e α_2 em relação à base ordenada $\{\beta_1, \beta_2\}$ de W?

5. Sejam $\alpha = (x_1, x_2)$ e $\beta = (y_1, y_2)$ vetores de R^2 tais que

$$x_1 y_1 + x_2 y_2 = 0, \qquad x_1^2 + x_2^2 = y_1^2 + y_2^2 = 1.$$

Demonstrar que $\mathscr{B} = \{\alpha, \beta\}$ é uma base de R^2. Determinar as coordenadas do vetor (a, b) em relação à base ordenada $\mathscr{B} = \{\alpha, \beta\}$. (As condições sobre α e β dizem, geometricamente, que α e β são perpendiculares e cada um tem comprimento 1.)

6. Seja V o espaço vetorial sobre o corpo dos números complexos das funções de R em C, isto é, o espaço das funções definidas sobre a reta real e tomando valores complexos. Sejam $f_1(x) = 1$, $f_2(x) = e^{ix}$, $f_3(x) = e^{-ix}$.

(a) Demonstrar que f_1, f_2 e f_3 são linearmente independentes.

(b) Sejam $g_1(x) = 1$, $g_2(x) = \cos x$, $g_3(x) = \operatorname{sen} x$. Determinar uma 3×3 matriz P inversível tal que

$$g_j = \sum_{i=1}^{3} P_{ij} f_i.$$

7. Seja V o espaço vetorial (real) das funções polinomiais de R em R de grau menor ou igual a 2, isto é, o espaço das funções f da forma

$$f(x) = c_0 + c_1 x + c_2 x^2.$$

Seja t um número real fixo e definamos

$$g_1(x) = 1,\ g_2(x) = x + t,\ g_3(x) = (x + t)^2.$$

Demonstrar que $\mathscr{B} = \{g_1, g_2, g_3\}$ é uma base de V. Se

$$f(x) = c_0 + c_1 x + c_2 x^2$$

quais são as coordenadas de f em relação a esta base ordenada $\mathscr{B}$?

2.5 Resumo de Linha-equivalência

Nesta seção utilizaremos alguns fatos elementares sobre bases e dimensão de espaços vetoriais de dimensão finita para completar nossa discussão de linha-equivalência de matrizes. Lembramos que se A é uma $m \times n$ matriz sobre o corpo F, os vetores-linhas de A são os vetores $\alpha_1, \ldots, \alpha_m$ em F^n definidos por

$$\alpha_i = (A_{i1}, \ldots, A_{in})$$

e que o espaço-linha de A é o subespaço de F^n gerado por estes vetores. O **posto-linha** de A é a dimensão do espaço-linha de A.

Se P é uma $k \times m$ matriz sobre F, então o produto $B = PA$ é uma $k \times m$ matriz cujos vetores-linhas $\beta_1, \ldots, \beta_k$ são combinações lineares

$$\beta_i = P_{i1}\alpha_1 + \ldots + P_{im}\alpha_m$$

dos vetores-linhas de A. Portanto, o espaço-linha de B é um subespaço do espaço-linha de A. Se P é uma $m \times m$ matriz inversível, então B é linha-equivalente a A de modo que a simetria da linha-equivalência, ou a equação $A = P^{-1}B$, implica que o espaço-linha de A também é um subespaço do espaço-linha de B.

Teorema 9. *Matrizes linha-equivalentes possuem o mesmo espaço-linha.*

Vemos assim que para estudar o espaço-linha de A podemos estudar o espaço-linha de uma matriz linha-reduzida à forma em escada que seja linha-equivalente a A. É o que passamos a fazer.

Teorema 10. *Seja* R *uma matriz não-nula linha-reduzida à forma em escada. Então os vetores-linhas não-nulos de* R *formam uma base do espaço-linha de* R.

Demonstração. Sejam $\rho_1, \ldots, \rho_r$ os vetores-linhas não-nulos de R:

$$\rho_i = (R_{i1}, \ldots, R_{in}).$$

Esses vetores certamente geram o espaço-linha de R; precisamos apenas demonstrar que eles são linearmente independentes. Como R é uma matriz linha-reduzida à forma em escada, existem inteiros positivos $k_1, \ldots, k_r$ tais que, para $i \leq r$, tem-se

$$\text{(2-18)} \qquad \begin{aligned} &\text{(a)} \quad R(i,j) = 0, \quad \text{se} \quad j < k_i \\ &\text{(b)} \quad R(i, k_j) = \delta_{ij} \\ &\text{(c)} \quad k_i < \ldots < k_r. \end{aligned}$$

Suponhamos que $\beta = (b_1, \ldots, b_n)$ seja um vetor do espaço-linha de R:

$$\text{(2-19)} \qquad \beta = c_1\rho_1 + \ldots + c_r\rho_r.$$

Afirmamos então que $c_j = b_{kj}$. De fato, por (2-18)

$$\text{(2-20)} \qquad \begin{aligned} b_{kj} &= \sum_{i=1}^{r} c_i R(i, k_j) \\ &= \sum_{i=1}^{r} c_i \delta_{ij} \\ &= c_j. \end{aligned}$$

Em particular, se $\beta = 0$, isto é, se $c_1\rho_1 + \ldots + c_r\rho_r = 0$, então c_j é necessariamente a k_j-ésima coordenada do vetor nulo, de modo que $c_j = 0$ para $j = 1, \ldots, r$. Assim, $\rho_1, \ldots, \rho_r$ são linearmente independentes.

Teorema 11. *Sejam* m *e* n *inteiros positivos e seja* F *um corpo. Suponhamos que* W *seja um subespaço de* F^n *e que* dim W $\leq$ m. *Então, existe exatamente uma* m $\times$ n *matriz sobre* F, *linha-reduzida à forma em escada, cujo espaço-linha é* W.

Demonstração. Existe pelo menos uma $m \times n$ matriz linha-reduzida à forma em escada cujo espaço-linha é W. Como dim $W \leq m$, podemos tomar m vetores $\alpha_1, \ldots, \alpha_m$ em W que geram W. Seja A a $m \times n$ matriz com vetores-linhas $\alpha_1, \ldots, \alpha_m$ e seja R uma matriz

linha-reduzida à forma em escada e linha-equivalente a A. Então, o espaço-linha de R é W.

Seja agora R uma matriz arbitrária, linha-reduzida à forma em escada e com espaço-linha W. Sejam $\rho_1, \ldots, \rho_r$ vetores-linha, não nulos, de R e suponhamos que o primeiro elemento não nulo de ρ_i esteja na coluna k_i, $i = 1, \ldots, r$. Os vetores $\rho_1, \ldots, \rho_r$ formam uma base W. Na demonstração do Teorema 10 observamos que se $\beta = (b_1, \ldots, b_n)$ estiver em W, então

$$\beta = c_1\rho_1 + \ldots + c_r\rho_r,$$

e $c_i = b_{k_i}$; em outras palavras, a única expressão de β como uma combinação linear de $\rho_1, \ldots, \rho_r$ é

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i. \tag{2-21}$$

Assim, todo vetor β está determinado se se conhecem as coordenadas b_{k_i}, $i = 1, \ldots, r$. Por exemplo, ρ_s é o único vetor em W cuja k-ésima coordenada é 1 e cuja k_i-ésima é 0 para $i \neq s$.

Suponhamos que β esteja em W e $\beta \neq 0$. Afirmamos que a primeira coordenada não-nula de β ocorre em uma das colunas k_s. Como

$$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i$$

e $\beta \neq 0$, podemos escrever

$$\beta = \sum_{j=s}^{r} b_{k_i}\rho_i, \qquad b_{k_s} \neq 0. \tag{2-22}$$

Das condições (2-18) tem-se que $R_{ij} = 0$ se $i > s$ e $j \leq k_s$. Portanto,

$$\beta = (0, \ldots, 0, b_{k_s}, \ldots, b_n), \; b_{k_s} \neq 0$$

e a primeira coordenada não-nula de β ocorre na coluna k_s. Notemos também que para cada k_s, $s = 1, \ldots, r$, existe um vetor em W cuja primeira coordenada não-nula ocorre na coluna k_s, a saber, o vetor ρ_s.

Está claro agora que R é determinada de modo único por W.

A descrição de R em termos de W será feita como segue. Consideremos todos os vetores $\beta = (b_1, \ldots, b_n)$ em W. Se $\beta \neq 0$, então a primeira coordenada não-nula de β deve ocorrer em uma certa coluna t:

$$\beta = (0, \ldots, 0, b_t, \ldots, b_n), \qquad b_t \neq 0$$

Sejam $k_1, \ldots, k_r$ os inteiros positivos t tais que exista algum $\beta \neq 0$ em W, cuja primeira coordenada não-nula ocorra na coluna t. Coloquemos os $k_1, \ldots, k_r$ na ordem $k_1 < k_2 < \ldots < k_r$. Para cada inteiro positivo k_s existirá um, e somente um, vetor ρ_s em W tal que a k_s-ésima coordenada de ρ_s seja 1 e a k_i-ésima coordenada de ρ_s seja 0 para $i \neq s$. Então, R é a $m \times n$ matriz cujos vetores-linhas são $\rho_1, \ldots, \rho_r, 0, \ldots, 0$.

Corolário. *Cada* m × n *matriz* A *é linha-equivalente a exatamente uma matriz linha-reduzida à forma em escada.*

Demonstração. Sabemos que A é linha-equivalente a pelo menos uma matriz R linha-reduzida à forma em escada. Se A é linha-equivalente a uma outra tal matriz R', então R é linha-equivalente a R'; logo, R e R' possuem o mesmo espaço-linha e são necessariamente idênticas.

Corolário. *Sejam* A *e* B m × n *matrizes sobre o corpo* F. *Então* A *e* B *são linha-equivalentes se, e somente se, possuem o mesmo espaço-linha.*

Demonstração. Sabemos que se A e B são linha-equivalentes, então possuem o mesmo espaço-linha. Suponhamos então que A e B possuam o mesmo espaço-linha. Ora, A é linha-equivalente a uma matriz R linha-reduzida à forma em escada e B é linha-equivalente a uma matriz R' linha-reduzida à forma em escada. Como A e B têm o mesmo espaço-linha, R e R' têm o mesmo espaço-linha. Portanto, $R = R'$ e A é linha-equivalente a B.

Resumindo — se A e B são $m \times n$ matrizes sobre o corpo F, as seguintes afirmações são equivalentes:

1. A e B são linha-equivalentes.
2. A e B possuem o mesmo espaço-linha.
3. $B = PA$, sendo P uma $m \times m$ matriz inversível.

Uma quarta afirmação equivalente é que os sistemas homogêneos $AX = 0$ e $BX = 0$ têm as mesmas soluções; contudo, apesar

de sabermos que a linha-equivalência de A e B implica que esses sistemas têm as mesmas soluções, parece melhor deixar a demonstração da recíproca para mais tarde.

2.6 Cálculos Concernentes a Subespaços

Gostaríamos agora de mostrar como as operações elementares fornecem um método padronizado de responder a certas perguntas concretas concernentes a subespaços de F^n. Já deduzimos os fatos de que precisaremos. Eles são aqui reunidos para conveniência do leitor. A discussão aplica-se a qualquer espaço vetorial n-dimensional sobre o corpo F se se toma uma base ordenada $\mathscr{B}$, fixa, e se descreve cada vetor α de V pela n-upla $(x_1, \ldots, x_n)$ que dá as coordenadas de α em relação à base ordenada $\mathscr{B}$.

Suponhamos que nos sejam dados m vetores $\alpha_1, \ldots, \alpha_m$ em F^n. Consideremos as seguintes perguntas:

1. Como se pode saber se os vetores $\alpha_1, \ldots, \alpha_m$ são linearmente dependentes? De maneira mais geral, como se determina a dimensão do subespaço W gerado por estes vetores?

2. Dado β em F^n, como se pode saber se β é uma combinação linear de $\alpha_1, \ldots, \alpha_m$, isto é, se β está no subespaço W?

3. Como se pode fazer uma descrição explícita do subespaço W?

A terceira pergunta é um tanto vaga, pois não especifica o que se quer dizer com uma "descrição explícita"; no entanto, esclareceremos este ponto fazendo o tipo de descrição que temos em mente. Com esta descrição, as perguntas (1) e (2) podem ser respondidas imediatamente.

Seja A a $m \times n$ matriz com vetores-linhas α_i:

$$\alpha_i = (A_{i1}, \ldots, A_{in}).$$

Efetuemos uma seqüência de operações elementares sobre linhas, começando com A e terminando com uma matriz R linha-reduzida à forma em escada. Já explicamos anteriormente como fazer isto. Neste ponto, a dimensão de W (o espaço-linha de A) é evidente, pois esta dimensão é simplesmente o número de vetores-linhas não-nulos de R. Se $\rho_1, \ldots, \rho_r$ são os vetores-linhas não-nulos de R, então $\mathscr{B} = \{\rho_1, \ldots, \rho_r\}$ é uma base de W. Se a primeira coordenada não-nula de ρ_i é a k_i-ésima, então temos, para $i \leq r$,

(a) $$R(i, j) = 0, \quad \text{se} \quad j < k_i$$

(b) $$R(i, k_j) = \delta_{ij}$$
(c) $$k_1 < \ldots < k_r.$$

O subespaço W consiste de todos os vetores

$$\beta = c_1\rho_1 + \ldots + c_r\rho_r$$

$$= \sum_{i=1}^{r} c_i(R_{i1}, \ldots, R_{in}).$$

As coordenadas $b_1, \ldots, b_n$ de um tal vetor β são então

(2-23) $$b_j = \sum_{i=1}^{r} c_i R_{ij}.$$

Em particular, $b_{k_j} = c_j$, e, se $\beta = (b_1, \ldots, b_n)$ é uma combinação linear dos ρ_i, então tem de ser a particular combinação linear

(2-24) $$\beta = \sum_{i=1}^{r} b_{k_i}\rho_i.$$

As condições sobre β para que (2-24) valha são:

(2-25) $$b_j = \sum_{i=1}^{r} b_{k_i} R_{ij}, \qquad j = 1, \ldots, n.$$

Ora, (2-25) é a descrição explícita do subespaço W gerado por $\alpha_1, \ldots, \alpha_m$, isto é, o subespaço consiste dos vetores β em F^n cujas coordenadas satisfazem (2-25). Que tipo de descrição é (2-25)? Em primeiro lugar, ela descreve W como o conjunto das soluções $\beta = (b_1, \ldots, b_n)$ do sistema (2-25) de equações lineares homogêneas. Este sistema de equações é de uma natureza muito particular, pois $(n - r)$ das coordenadas como combinações lineares das r coordenadas distinguidas $b_{k_1}, \ldots, b_{k_r}$. Tem-se completa liberdade de escolha das coordenadas b_{k_i}, isto é, se $c_1, \ldots, c_r$ são r escalares *arbitrários*, existe um e somente um vetor β em W cuja k_i-ésima coordenada é c_i.

O ponto importante aqui é o seguinte: Dados os vetores α_i, a linha-redução é um método direto de determinar os inteiros $r, k_1, \ldots, k_r$ e os escalares R_{ij} que dão a descrição (2-25) do subespaço gerado por $\alpha_1, \ldots, \alpha_m$. Deve-se observar, como no Teore-

ma 11, que todo subespaço W de F^n possui uma descrição do tipo (2-25). Devemos destacar também algumas coisas sobre a pergunta (2). Já mostramos na seção 1.4 como se pode encontrar uma $m \times m$ matriz inversível P tal que $R = PA$. O conhecimento de P nos permite determinar os escalares $x_1, \ldots, x_m$ tais que

$$\beta = x_1\alpha_1 + \ldots + x_m\alpha_m$$

quando isto é possível. Como os vetores-linhas de R são dados por

$$\rho_i = \sum_{j=1}^{m} P_{ij}\alpha_j$$

temos que se β é uma combinação linear dos α_j, então

$$\begin{aligned}\beta &= \sum_{i=1}^{r} b_{k_i}\rho_i \\ &= \sum_{i=1}^{r} b_{k_i} \sum_{j=1}^{m} P_{ij}\alpha_j \\ &= \sum_{j=1}^{m} \sum_{i=1}^{r} b_{k_i}P_{ij}\alpha_j\end{aligned}$$

e portanto

$$x_j = \sum_{i=1}^{r} b_{k_i}P_{ij}$$

é uma escolha possível para os x_j (podem existir muitas).

O problema de saber se $\beta = (b_1, \ldots, b_n)$ é uma combinação linear dos α_i, e, em caso afirmativo, quais são os escalares x_i, pode também ser considerado perguntando-se se o sistema de equações

$$\sum_{i=1}^{m} A_{ij}x_i = b_j, \quad j = 1, \ldots, n$$

admite soluções e quais são elas. A matriz dos coeficientes deste sistema de equações é a $n \times$ m matriz B com vetores-*colunas* $\alpha_1, \ldots, \alpha_m$. No Capítulo 1 discutimos o uso de operações elementares sobre linhas na resolução de um sistema de equações $BX = Y$. Consideremos um exemplo no qual adotamos este último ponto de vista ao respondermos a perguntas sobre subespaços de F^n.

Exemplo 21. Proponhamos o seguinte problema: seja W o subespaço de R^4 gerado pelos vetores

$$\begin{aligned} \alpha_1 &= (1,\ 2,\ 2,\ 1) \\ \alpha_2 &= (0,\ 2,\ 0,\ 1) \\ \alpha_3 &= (-2,\ 0,\ -4,\ 3). \end{aligned}$$

(a) Demonstrar que $\alpha_1, \alpha_2, \alpha_3$ formam uma base de W, isto é, que estes vetores são linearmente independentes.

(b) Seja $\beta = (b_1, b_2, b_3, b_4)$ um vetor em W. Quais são as coordenadas de β em relação à base ordenada $\{\alpha_1, \alpha_2, \alpha_3\}$?

(c) Sejam

$$\begin{aligned} \alpha'_1 &= (1,\ 0,\ 2,\ 0) \\ \alpha'_2 &= (0,\ 2,\ 0,\ 1) \\ \alpha'_3 &= (0,\ 0,\ 0,\ 3). \end{aligned}$$

Mostrar que $\alpha'_1, \alpha'_2, \alpha'_3$ formam uma base de W.

(d) Se β está em W, indiquemos por X a matriz das coordenadas de β em relação à base dos α e por X' a matriz das coordenadas de β em relação à base dos α'. Determinar a 3×3 matriz P tal que $X = PX'$ para todo β em W.

Para responder a estas questões usando o primeiro método, formamos a matriz A com vetores-linha $\alpha_1, \alpha_2, \alpha_3$, determinamos a matriz R, linha-reduzida à forma em escada e equivalente a A e, simultaneamente, efetuamos as mesmas operações na matriz unidade para obter a matriz inversível Q tal que $R = QA$:

$$\begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix} \rightarrow R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \rightarrow Q = \tfrac{1}{6}\begin{bmatrix} 6 & -6 & 0 \\ -2 & 5 & -1 \\ 4 & -4 & 2 \end{bmatrix}$$

(a) É claro que R possui posto 3 e portanto α_1, α_2 e α_3 são independentes.

(b) Quais vetores $\beta = (b_1, b_2, b_3, b_4)$ estão em W? Temos a base de W formada pelos vetores-linha de R, ρ_1, ρ_2, ρ_3. Podemos

ver imediatamente que o espaço gerado por ρ_1, ρ_2, ρ_3 é formado pelos vetores β para os quais $b_3 = 2b_1$. Para um tal β temos

$$\begin{aligned}\beta &= b_1\rho_1 + b_2\rho_2 + b_4\rho_4\\ &= [b_1, b_2, b_4]\, R\\ &= [b_1, b_2, b_4]\, QA\\ &= x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3\end{aligned}$$

onde $x_i = [b_1, b_2, b_4]\, Q_i$:

(2-26)
$$\begin{aligned}x_1 &= b_1 - \tfrac{1}{3}b_2 + \tfrac{2}{3}b_4\\ x_2 &= -b_1 + \tfrac{5}{6}b_2 - \tfrac{2}{3}b_4\\ x_3 &= - \tfrac{1}{6}b_2 + \tfrac{1}{3}b_4\end{aligned}$$

(c) Os vetores α'_1, α'_2, α'_3 são todos da forma (y_1, y_2, y_3, y_4) com $y_3 = 2y_1$ e, portanto, estão em W. Pode-se ver facilmente que eles são independentes.

(d) A matriz P tem para colunas

$$P_j = [\alpha'_j]_{\mathscr{B}}$$

onde $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3\}$. As equações (2-26) nos dizem como determinar as matrizes das coordenadas de α'_1, α'_2, α'_3. Por exemplo com $\beta = \alpha'_1$ temos $b_1 = 1$, $b_2 = 0$, $b_3 = 2$, $b_4 = 0$ e

$$\begin{aligned}x_1 &= 1 - \tfrac{1}{3}(0) + \tfrac{2}{3}(0) = 1\\ x_2 &= -1 + \tfrac{5}{6}(0) - \tfrac{2}{3}(0) = -1\\ x_3 &= - \tfrac{1}{6}(0) + \tfrac{1}{3}(0) = 0.\end{aligned}$$

Assim, $\alpha'_1 = \alpha_1 - \alpha_2$. Analogamente, obtemos $\alpha'_2 = \alpha_2$ e $\alpha'_3 = 2\alpha_1 - - 2\alpha_2 + \alpha_3$. Logo

$$P = \begin{bmatrix} 1 & 0 & 2\\ -1 & 1 & -2\\ 0 & 0 & 1\end{bmatrix}.$$

Vejamos, agora, como responderíamos às questões usando o segundo método. Formemos a 4×3 matriz B com vetores-colunas $\alpha_1, \alpha_2, \alpha_3$:

$$B = \begin{bmatrix} 1 & 0 & -2\\ 2 & 2 & 0\\ 2 & 0 & -4\\ 1 & 1 & 3\end{bmatrix}.$$

Perguntamos para quais y_1, y_2, y_3, y_4 o sistema $BX = Y$ admite solução.

$$\begin{bmatrix} 1 & 0 & -2 & y_1 \\ 2 & 2 & 0 & y_2 \\ 2 & 0 & -4 & y_3 \\ 1 & 1 & 3 & y_4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -2 & y_1 \\ 0 & 2 & 4 & y_2 - 2y_1 \\ 0 & 0 & 0 & y_3 - 2y_1 \\ 0 & 1 & 5 & y_4 - y_1 \end{bmatrix} \to$$

$$\begin{bmatrix} 1 & 0 & -2 & y_1 \\ 0 & 0 & -6 & y_2 - 2y_4 \\ 0 & 1 & 5 & y_4 - y_1 \\ 0 & 0 & 0 & y_3 - 2y_1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & y_1 - \frac{1}{3}y_2 + \frac{2}{3}y_4 \\ 0 & 0 & 1 & \frac{1}{6}(2y_4 - y_2) \\ 0 & 1 & 0 & -y_1 + \frac{5}{6}y_2 - \frac{2}{3}y_4 \\ 0 & 0 & 0 & y_3 - 2y_1 \end{bmatrix}$$

Portanto, a condição para que o sistema $BX = Y$ admita uma solução é $y_3 = 2y_1$. Então, $\beta = (b_1, b_2, b_3, b_4)$ está em W se, e somente se, $b_3 = 2b_1$. Se β está em W, então as coordenadas (x_1, x_2, x_3, x_4) na base ordenada $\{\alpha_1, \alpha_2, \alpha_3\}$ podem ser obtidas da última matriz acima. Para essas coordenadas, obtemos novamente as fórmulas (2-26).

As questões (c) e (d) permanecem com a resolução anterior.

Exemplo 22. Consideremos a 5×5 matriz

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

e os seguintes problemas a respeito de A.

(a) Determinar uma matriz inversível P tal que PA seja uma matriz R, linha-reduzida à forma em escada.

(b) Determinar uma base para o espaço-linha W de A.

(c) Dizer quais vetores $(b_1, b_2, b_3, b_4, b_5)$ estão em W.

(d) Na base escolhida em (*b*), determinar a matriz das coordenadas de cada vetor $(b_1, b_2, b_3, b_4, b_5)$ de W.

(e) Escrever cada vetor $(b_1, b_2, b_3, b_4, b_5)$ de W como combinação linear das linhas de A.

(f) Descrever explicitamente o espaço vetorial V das 5×1 matrizes-coluna X tais que $AX = 0$.

(g) Determinar uma base de V.

(h) Para quais 5×1 matrizes-coluna Y a equação $AX = Y$ possui soluções X?

Para resolver estes problemas, formamos a matriz completa A' do sistema $AX = Y$ e aplicamos uma seqüência apropriada de operações sobre as linhas de A':

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 1 & 2 & -1 & -1 & 0 & y_2 \\ 0 & 0 & 1 & 4 & 0 & y_3 \\ 2 & 4 & 1 & 10 & 1 & y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & -1 & -4 & 0 & -y_1 + y_2 \\ 0 & 0 & 1 & 4 & 0 & y_3 \\ 0 & 0 & 1 & 4 & 1 & -2y_1 + y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & 1 & 4 & 0 & y_1 - y_2 \\ 0 & 0 & 0 & 0 & 0 & -y_1 + y_2 + y_3 \\ 0 & 0 & 0 & 0 & 1 & -3y_1 + y_2 + y_4 \\ 0 & 0 & 0 & 0 & 1 & y_5 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & y_1 \\ 0 & 0 & 1 & 4 & 0 & y_1 - y_2 \\ 0 & 0 & 0 & 0 & 1 & y_5 \\ 0 & 0 & 0 & 0 & 0 & -y_1 + y_2 + y_3 \\ 0 & 0 & 0 & 0 & 0 & -3y_1 + y_2 + y_4 - y_5 \end{bmatrix}$$

(a) Se

$$PY = \begin{bmatrix} y_1 \\ y_1 - y_2 \\ y_5 \\ -y_1 + y_2 + y_3 \\ -3y_1 + y_2 + y_4 - y_5 \end{bmatrix}$$

para todo Y, então

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix}$$

logo PA é a matriz linha-reduzida à forma em escada

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot$$

Deve ser ressaltado que a matriz P não é única. Existem, de fato, muitas matrizes inversíveis P (que surgem das diferentes escolhas das operações usadas para reduzir A') tais que $PA = R$.

(b) Como base de W podemos escolher as linhas não nulas

$$\begin{aligned} \rho_1 &= (1 \quad 2 \quad 0 \quad 3 \quad 0) \\ \rho_2 &= (0 \quad 0 \quad 1 \quad 4 \quad 0) \\ \rho_3 &= (0 \quad 0 \quad 0 \quad 0 \quad 1) \end{aligned}$$

de R.

(c) O espaço-linha W é constituído de todos os vetores da forma

$$\begin{aligned} \beta &= c_1\rho_1 + c_2\rho_2 + c_3\rho_3 \\ &= (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3) \end{aligned}$$

onde c_1, c_2, c_3 são escalares arbitrários. Assim, $(b_1, b_2, b_3, b_4, b_5)$ está em W se, e somente se,

$$(b_1, b_2, b_3, b_4, b_5) = b_1\rho_1 + b_3\rho_2 + b_5\rho_3$$

o que acontece se, e somente se,

$$\begin{aligned} b_2 &= 2b_1 \\ b_4 &= 3b_1 + 4b_3. \end{aligned}$$

Estas equações são exemplos do sistema geral (2-25) e nos permitem dizer imediatamente se um dado vetor está em W. Assim, $(-5, -10, 1, -11, 20)$ é uma combinação das linhas de A, o mesmo não acontecendo com $(1, 2, 3, 4, 5)$.

(d) A matriz das coordenadas do vetor $(b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5)$ na base $\{\rho_1, \rho_2, \rho_3\}$ é, evidentemente,

$$\begin{bmatrix} b_1 \\ b_3 \\ b_5 \end{bmatrix}.$$

(e) Existem vários modos de escrever os vetores de W como combinações lineares das linhas de A. O método mais simples talvez seja o de seguir o primeiro processo, indicado antes do Exemplo 21:

$$\begin{aligned} \beta &= (b_1, 2b_1, b_3, 3b_1 + 4b_3, b_5) \\ &= [b_1, b_3, b_5, 0, 0] \cdot R \\ &= [b_1, b_3, b_5, 0, 0] \cdot PA \end{aligned}$$

$$= [b_1, b_3, b_5, 0, 0] \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 \\ -3 & 1 & 0 & 1 & -1 \end{bmatrix} \cdot A$$

$$= [b_1 + b_3, -b_3, 0, 0, b_5] \cdot A.$$

Em particular, com $\beta = (-5, -10, 1, -11, 20)$, temos

$$\beta = (-4, -1, 0, 0, 20) \begin{bmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(f) As equações no sistema $RX = 0$ são

$$\begin{aligned} x_1 + 2x_2 + 3x_4 &= 0 \\ x_3 + 4x_4 &= 0 \\ x_5 &= 0. \end{aligned}$$

Assim, V é constituído de todas as colunas da forma

$$X = \begin{bmatrix} -2x_2 - 3x_4 \\ x_2 \\ -4x_4 \\ x_4 \\ 0 \end{bmatrix}$$

onde x_2 e x_4 são arbitrários.

(g) As colunas

$$\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}$$

formam uma base de V. Este é um exemplo de uma base como a descrita no Exemplo 15.

(h) A equação $AX = Y$ tem soluções se, e somente se,

$$\begin{aligned} -\ y_1 + y_2 + y_3 \qquad &= 0 \\ -\ 3y_1 + y_2 + y_4 - y_5 &= 0. \end{aligned}$$

Exercícios

1. Seja $s < n$ e A uma $s \times n$ matriz com elementos no corpo F. Usar o Teorema 4 (não sua demonstração) para mostrar que existe um X, não nulo, em $F^{n \times 1}$, tal que $AX = 0$.

2. Sejam

$$\alpha_1 = (1, 1, -2, 1), \quad \alpha_2 = (3, 0, 4, -1), \quad \alpha_3 = (-1, 2, 5, 2).$$

Sejam

$$\alpha = (4, -5, 9, -7), \quad \beta = (3, 1, -4, 4), \quad \gamma = (-1, 1, 0, 1).$$

(a) Quais dos vetores α, β, γ estão no subespaço de R^4 gerado pelos α_i?
(b) Quais dos vetores α, β, γ estão no subespaço de C^4 gerado pelos α_i?
(c) Isto sugere algum teorema?

3. Consideremos os vetores em R^4 definidos por

$$\alpha_1 = (-1, 0, 1, 2), \quad \alpha_2 = (3, 4, -2, 5), \quad \alpha_3 = (1, 4, 0, 9).$$

Determinar um sistema de equações lineares homogêneas para o qual o espaço das soluções seja exatamente o subespaço de R^4 gerado pelos três vetores dados.

4. Em C^3 sejam

$$\alpha_1 = (1,\ 0,\ -i), \quad \alpha_2 = (1 + i,\ 1 - i,\ 1), \quad \alpha_3 = (i,\ i,\ i).$$

Demonstrar que estes vetores formam uma base de C^3. Quais são as coordenadas do vetor (a, b, c) em relação a esta base?

5. Fazer uma descrição explícita do tipo (2-25) para os vetores

$$\beta = (b_1,\ b_2,\ b_3,\ b_4,\ b_5)$$

em R^5 que são combinações lineares dos vetores

$$\alpha_1 = (1,\ 0,\ 2,\ 1,\ -1), \qquad \alpha_2 = (-1,\ 2,\ -4,\ 2,\ 0),$$

$$\alpha_3 = (2,\ -1,\ 5,\ 2,\ 1), \quad \text{e} \quad \alpha_4 = (2,\ 1,\ 3,\ 5,\ 2).$$

6. Seja V o espaço vetorial real gerado pelas linhas da matriz

$$A = \begin{bmatrix} 3 & 21 & 0 & 9 & 0 \\ 1 & 7 & -1 & -2 & -1 \\ 2 & 14 & 0 & 6 & 1 \\ 6 & 42 & -1 & 13 & 0 \end{bmatrix}.$$

(a) Determinar uma base de V.

(b) Dizer quais vetores $(x_1, x_2, x_3, x_4, x_5)$ são elementos de V.

(c) Se $(x_1, x_2, x_3, x_4, x_5)$ estiver em V, quais serão as suas coordenadas em relação à base escolhida na parte (a)?

7. Seja A uma $m \times n$ matriz sobre o corpo F e consideremos o sistema de equações $AX = Y$. Demonstrar que este sistema de equações admite uma solução se, e somente se, o posto-linha de A é igual ao posto-linha da matriz completa do sistema.

Capítulo 3
TRANSFORMAÇÕES LINEARES

3.1 Transformações Lineares

Introduziremos agora as transformações lineares, objetos que estudaremos na maior parte do restante deste livro. O leitor poderá achar útil ler (ou reler) a discussão sobre funções no Apêndice, visto que usaremos livremente a terminologia daquela discussão.

Definição. *Sejam* V *e* W *espaços vetoriais sobre o corpo* F. *Uma* **transformação linear** *de* V *em* W *é uma função* T *de* V *em* W *tal que*

$$T(c\alpha + \beta) = c(T\alpha) + T\beta$$

para todos α *e* β *em* V *e todos escalares* c *em* F.

Exemplo 1. Se V é um espaço vetorial arbitrário, a transformação idêntica I, definida por $I\alpha = \alpha$, é uma transformação linear de V em V. A **transformação nula** 0, definida por $0\alpha = 0$, é uma transformação linear de V em V.

Exemplo 2. Seja F um corpo e seja V o espaço das funções polinomiais f de F em F, dadas por

$$f(x) = c_0 + c_1 x + \ldots + c_k x^k.$$

Seja

$$(Df)(x) = c_1 + 2c_2 x + \ldots + k c_k x^{k-1}.$$

Então, D é uma transformação linear de V em V — a transformação derivação.

Exemplo 3. Seja A uma $m \times n$ matriz, fixa, com elementos no corpo F. A função T definida por $T(X) = AX$ é uma transformação linear de $F^{n \times 1}$ em $F^{m \times 1}$. A função U definida por $U(\alpha) = \alpha A$ é uma transformação linear de F^m em F^n.

Exemplo 4. Seja P uma $m \times m$ matriz fixa com elementos no corpo F e seja Q uma $n \times n$ matriz fixa sobre F. Definamos uma função T do espaço $F^{m \times n}$ em si mesmo por $T(A) = PAQ$. Então T é uma transformação linear de $F^{m \times n}$ em $F^{m \times n}$, pois

$$\begin{aligned} T(cA + B) &= P(cA + B)Q \\ &= (cPA + PB)Q \\ &= cPAQ + PBQ \\ &= cT(A) + T(B). \end{aligned}$$

Exemplo 5. Seja R o corpo dos números reais e seja V o espaço das funções de R em R que são *contínuas*. Definamos T por

$$(Tf)(x) = \int_0^x f(t)\, dt.$$

Então T é uma transformação linear de V em V. A função Tf, além de contínua, possui a primeira derivada contínua. A linearidade da integração é uma de suas propriedades fundamentais.

O leitor não deverá encontrar nenhuma dificuldade para verificar que as transformações definidas nos Exemplos 1, 2, 3 e 5 são transformações lineares. Ampliaremos consideravelmente nossa lista de exemplos à medida que aprendermos mais sobre transformações lineares.

É importante notar que se T é uma transformação linear de V em W, então $T(0) = 0$; pode-se ver isto a partir da definição, pois

$$T(0) = T(0 + 0) = T(0) + T(0).$$

Este ponto freqüentemente causa confusão à pessoa que estuda álgebra linear pela primeira vez, desde que ela provavelmente tomou contato com uma utilização ligeiramente diferente do termo "função linear". Um comentário breve deverá eliminar a confusão. Suponhamos que V seja o espaço vetorial R^1. Uma transformação linear de V em V é então um tipo especial de função com valores reais definida sobre a reta real R. Num curso de cálculo, ter-se-ia

provavelmente a denominação linear para uma tal função se seu gráfico fosse uma reta. Uma transformação linear de R^1 em R^1, de acordo com nossa definição, será uma função de R em R, cujo gráfico é uma reta *que passa pela origem.*

Além da propriedade $T(0) = 0$, destaquemos outra propriedade de uma transformação linear arbitrária T. Uma tal transformação "conserva" combinações lineares; isto é, se $\alpha_1, \ldots, \alpha_n$ são vetores em V e $c_1, \ldots, c_n$ são escalares, então

$$T(c_1\alpha_1 + \ldots + c_n\alpha_n) = c_1(T\alpha_1) + \ldots + c_n(T\alpha_n).$$

Isto decorre imediatamente da definição. Por exemplo,

$$\begin{aligned} T(c_1\alpha_1 + c_2\alpha_2) &= T(c_1\alpha_1) + T(c_2\alpha_2) \\ &= c_1(T\alpha_1) + c_2(T\alpha_2). \end{aligned}$$

Teorema 1. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F *e seja* $\{\alpha_1, \ldots, \alpha_n\}$ *uma base ordenada de* V. *Seja* W *um espaço vetorial sobre o mesmo corpo* F *e sejam* $\beta_1, \ldots, \beta_n$ *vetores arbitrários em* W. *Então, existe exatamente uma transformação linear* T *de* V *em* W *tal que*

$$T\alpha_j = \beta_j, \quad j = 1, \ldots, n.$$

Demonstração. Para demonstrar que existe pelo menos uma transformação linear T com $T\alpha_j = \beta_j$ procedemos como segue. Dado α em V, existe uma única n-upla $(x_1, \ldots, x_n)$ tal que

$$\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n.$$

Para este vetor α, definamos

$$T\alpha = x_1\beta_1 + \ldots + x_n\beta_n.$$

Então, T é uma regra bem definida para se associar a cada vetor α em V um vetor $T\alpha$ em W. Pela definição, é evidente que $T\alpha_j = \beta_j$ para todo j. Para ver que T é linear, seja

$$\beta = y_1\alpha_1 + \ldots + y_n\alpha_n$$

em V e c um escalar arbitrário. Ora,

$$c\alpha + \beta = (cx_1 + y_1)\alpha_1 + \ldots + (cx_n + y_n)\alpha_n$$

portanto, pela definição,

$$T(c\alpha + \beta) = (cx_1 + y_1)\beta_1 + \dots + (cx_n + y_n)\beta_n.$$

Por outro lado,

$$c(T\alpha) + T\beta = c \sum_{i=1}^{n} x_i\beta_i + \sum_{i=1}^{n} y_i\beta_i$$

$$= \sum_{i=1}^{n} (cx_i + y_i)\beta_i$$

e assim

$$T(c\alpha + \beta) = c(T\alpha) + T\beta.$$

Se U é uma transformação linear de V em W com $U\alpha_j = \beta_j$, $j = 1, \dots, n$, então, para o vetor $\alpha = \sum_{i=1}^{n} x_i\alpha_i$, temos

$$U\alpha = U\left(\sum_{i=1}^{n} x_i\alpha_i\right)$$

$$= \sum_{i=1}^{n} x_i(U\alpha_i)$$

$$= \sum_{i=1}^{n} x_i\beta_i$$

de modo que U é exatamente a regra T que definimos acima. Isto mostra que a transformação linear T com $T\alpha_j = \beta_j$ é única.

O Teorema 1 é bastante elementar; no entanto, ele é tão fundamental, que decidimos enunciá-lo formalmente. O conceito de função é muito geral. Se V e W são espaços vetoriais (não nulos), existe um grande número de funções de V em W. O Teorema 1 ajuda ressaltar o fato de que funções que sejam lineares são extremamente especiais.

Exemplo 6. Os vetores

$$\alpha_1 = (1, 2)$$

e

$$\alpha_2 = (3, 4)$$

são linearmente independentes e portanto constituem uma base de R^2. De acordo com o Teorema 1, existe uma única transformação linear de R^2 em R^3 tal que

$$T\alpha_1 = (3, 2, 1)$$

e

$$T\alpha_2 = (6, 5, 4)$$

Assim, devemos ser capazes de determinar $T(\varepsilon_1)$. Determinamos escalares c_1 e c_2 tais que $\varepsilon_1 = c_1\alpha_1 + c_2\alpha_2$ e sabemos que $T\varepsilon_1 = c_1 T\alpha_1 + c_2 T\alpha_2$. Se $(1, 0) = c_1(1, 2) + c_2(3, 4)$, então $c_1 = -2$ e $c_2 = 1$. Assim,

$$\begin{aligned} T(1, 0) &= -2(3, 2, 1) + (6, 5, 4) \\ &= (0, 1, 2). \end{aligned}$$

Exemplo 7. Seja T uma transformação linear do espaço F^m das m-uplas no espaço F^n das n-uplas. O Teorema 1 nos diz que T é determinado de modo único pela seqüência de vetores $\beta_1, \ldots, \beta_m$, onde

$$\beta_i = T\varepsilon_i \quad i = 1, \ldots, m$$

Em suma, T é determinada de modo único pelas imagens dos vetores da base canônica. A determinação é

$$\begin{aligned} \alpha &= (x_1, \ldots, x_m) \\ T\alpha &= x_1\beta_1 + \ldots + x_m\beta_m. \end{aligned}$$

Se B for a $m \times n$ matriz com vetores-linha $\beta_1, \ldots, \beta_m$ isto significa que

$$T\alpha = \alpha B.$$

Em outras palavras, se $\beta_i = (B_{i1}, \ldots, B_{in})$, então

$$T(x_1, \ldots, x_n) = [x_1, \ldots, x_m] \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & & \vdots \\ B_{m1} & \cdots & B_{mn} \end{bmatrix}.$$

Esta é uma descrição muita explícita da transformação linear. Na Seção 3.4 faremos um estudo mais profundo da relação entre trans-

formações lineares e matrizes. Não prosseguiremos com a descrição $T\alpha = \alpha\beta$ porque nela a matriz B está à direita do vetor α e isto pode causar alguma confusão. O objetivo deste exemplo é mostrar que podemos dar uma descrição explícita, e razoavelmente simples, de todas as transformações lineares de F^m em F^n.

Se T é uma transformação linear de V em W, a imagem de T não é apenas um subconjunto de W; é um subespaço de W. Seja R_T a imagem de T, isto é, o conjunto de todos os vetores β de W tais que $\beta = T\alpha$ para algum α em V. Sejam β_1 e β_2 elementos de R_T e seja c um escalar. Existem vetores α_1 e α_2 em V tais que $T(\alpha_1 = \beta_1$ e $T\alpha_2 = \beta_2$. Sendo T linear

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c\beta_1 + \beta_2 \end{aligned}$$

o que mostra que $c\beta_1 + \beta_2$ também está em R_T.

Um outro subespaço interessante, associado com a transformação linear T, é o conjunto N dos vetores α em V tais que $T\alpha = 0$. Este conjunto é um subespaço de V pois

(a) $T(0) = 0$, de modo que N é não vazio;

(b) se $T\alpha_1 = T\alpha_2 = 0$, então

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c0 + 0 \\ &= 0 \end{aligned}$$

de modo que $c\alpha_1 + \alpha_2$ está em N.

Definição. *Sejam* V *e* W *espaços vetoriais sobre o corpo* F *e seja* T *uma transformação linear de* V *em* W. *O* **núcleo** (*ou espaço nulo*) *de* T *é o conjunto de todos os vetores* α *de* V *tais que* $T\alpha = 0$.

Se V *é de dimensão finita, o* **posto** *de* T *é a dimensão da imagem de* T *e a* **nulidade** *de* T *é a dimensão do núcleo de* T.

O teorema seguinte é um dos resultados mais importantes na álgebra linear.

Teorema 2. *Sejam* V *e* W *espaços vetoriais sobre o corpo* F *e seja* T *uma transformação linear de* V *em* W. *Suponhamos que* V *seja de dimensão finita. Então*

$$\textit{posto}\,(\mathrm{T}) + \textit{nulidade}\,(\mathrm{T}) = \dim \mathrm{V}.$$

Demonstração. Seja $\{\alpha_1, \ldots, \alpha_k\}$ uma base de N, o núcleo de T. Existem vetores $\alpha_{k+1}, \ldots, \alpha_n$ em V tais que $\{\alpha_1, \ldots, \alpha_n\}$ seja uma base de V. Demonstraremos agora que $\{T\alpha_{k+1}, \ldots, T\alpha_n\}$ é uma base da imagem de T. Os vetores $T\alpha_1, \ldots, T\alpha_n$ certamente geram a imagem de T e, como $T\alpha_j = 0$, para $j \leq k$, vemos que $T\alpha_{n+1}, \ldots, T\alpha_n$ geram a imagem. Para ver que esses vetores são independentes, suponhamos que existam escalares c_i tais que

$$\sum_{i=k+1}^{n} c_i(T\alpha_i) = 0.$$

Isto diz que

$$T\left(\sum_{i=k+1}^{n} c_i\alpha_i\right) = 0$$

e, conseqüentemente, o vetor $\alpha = \sum\limits_{i=k+1}^{n} c_i\alpha_i$ está no núcleo de T. Como $\alpha_1, \ldots, \alpha_k$ formam uma base de N, existem, necessariamente, escalares $b_1, \ldots, b_k$ tais que

$$\alpha = \sum_{i=1}^{k} b_i\alpha_i.$$

Assim

$$\sum_{i=1}^{k} b_i\alpha_i - \sum_{j=k+1}^{n} c_j\alpha_j = 0$$

e, como $\alpha_1, \ldots, \alpha_n$ são linearmente independentes, devemos ter

$$b_1 = \ldots = b_k = c_{k+1} = \ldots = c_n = 0.$$

Se r é o posto de T, o fato de $T\alpha_{k+1}, \ldots, T\alpha_n$ formarem uma base da imagem de T nos diz que $r = n - k$. Como k é a nulidade de T e n é a dimensão de V, está completa a demonstração.

Teorema 3. *Se* A *é uma* m × n *matriz com elementos no corpo* F, *então*

$$\textit{posto-linha}\ (A) = \textit{posto-coluna}\ (A).$$

Demonstração. Seja T a transformação linear de $F^{n\times 1}$ em $F^{m\times 1}$ definida por $T(X) = AX$. O núcleo de T é o espaço solu-

ção do sistema $AX = 0$, isto é, o conjunto de todas as matrizes-coluna X tais que $AX = 0$. A imagem de T é o conjunto de todas as $m \times 1$ matrizes-coluna Y tais que $AX = Y$ tenha solução em X. Se $A_1, \ldots, A_n$ forem as colunas de A, então

$$AX = x_1A_1 + \ldots + x_nA_n$$

de modo que a imagem de T é o subespaço gerado pelas colunas de A. Em outras palavras, a imagem de T é o espaço-coluna de A. Portanto

$$\text{posto } (T) = \text{posto-coluna } (A).$$

O Teorema 2 nos diz que, sendo S o espaço solução do sistema $AX = 0$, temos

$$\dim S + \text{posto-coluna } (A) = n.$$

Voltando ao Exemplo 15 do Capítulo 2, vimos que se r for a dimensão do espaço-linha de A então o espaço solução S terá uma base contendo $n - r$ vetores:

$$\dim S = n - \text{posto-linha } (A).$$

É evidente agora que

$$\text{posto-linha } (A) = \text{posto-coluna } (A).$$

A demonstração que acabamos de dar do Teorema 3 depende de cálculos explícitos envolvendo sistemas de equações lineares. Existe uma demonstração mais conceitual que não depende de tais cálculos. Daremos esta demonstração na Seção 3.7.

Exercícios

1. Quais das seguintes funções T de R^2 em R^2 são transformações lineares?

(a) $T(x_1, x_2) = (1 + x_1, x_2)$;
(b) $T(x_1, x_2) = (x_2, x_1)$;
(c) $T(x_1, x_2) = (x_1^2, x_2)$;
(d) $T(x_1, x_2) = (\text{sen } x_1, x_2)$;
(e) $T(x_1, x_2) = (x_1 - x_2, 0)$.

2. Determinar a imagem, o posto, o núcleo e a nulidade da transformação nula e da transformação idêntica sobre um espaço V de dimensão finita.

3. Descrever a imagem e o núcleo da transformação derivação do Exemplo 2. Fazer o mesmo para a transformação integração do Exemplo 5.

4. Existe uma transformação linear T de R^3 em R^2 tal que $T(1, -1, 1) = (1, 0)$ e $T(1, 1, 1) = (0, 1)$?

5. Se

$$\begin{aligned} \alpha_1 &= (1, -1), & \beta_1 &= (1, 0) \\ \alpha_2 &= (2, -1), & \beta_2 &= (0, 1) \\ \alpha_3 &= (-3, 2), & \beta_3 &= (1, 1) \end{aligned}$$

existe uma transformação linear T de R^2 em R^2 tal que $T\alpha_i = \beta_i$ para $i = 1, 2$ e 3?

6. Descrever explicitamente (como no Exercício 1) a transformação linear T de F^2 em F^2 tal que $T\varepsilon_1 = (a, b)$, $T\varepsilon_2 = (c, d)$.

7. Seja F um subcorpo do corpo dos números complexos e seja T a função de F^3 em F^3 definida por

$$T(x_1, x_2, x_3) = (x_1 - x_2 + 2x_3, 2x_1 + x_2, -x_1 - 2x_2 + 2x_3).$$

(a) Verificar que T é uma transformação linear.
(b) Se (a, b, c) é um vetor em F^3, quais as condições sobre a, b, c, para que o vetor esteja na imagem de T? Qual é o posto de T?
(c) Quais são as condições sobre a, b e c para que o vetor esteja no núcleo de T? Qual é a nulidade de T?

8. Descrever explicitamente uma transformação linear de R^3 em R^3 cuja imagem seja o subespaço por $(1, 0, -1)$ e $(1, 2, 2)$.

9. Seja V o espaço vetorial das $n \times n$ matrizes sobre o corpo F e seja B uma $n \times n$ matriz fixa. Se

$$T(A) = AB - BA$$

verificar que T é uma transformação linear de V em V.

10. Seja V o conjunto dos números complexos considerado como um espaço vetorial sobre o corpo dos números *reais* (operações usuais). Determinar uma função de V em V que seja uma transformação linear sobre o espaço vetorial acima, mas que não seja uma transformação linear sobre C^1, isto é, que não seja linear complexa.

11. Seja V o espaço das $n \times 1$ matrizes sobre F e seja W o espaço das $m \times 1$ matrizes sobre F. Seja A uma $m \times n$ matriz fixa sobre F e seja T a transformação linear de V em W definida por $T(X) = AX$. Demonstrar que T é a transformação nula se, e somente se, A é a matriz nula.

12. Seja V um espaço vetorial n-dimensional sobre o corpo F e seja T uma transformação linear de V em V tal que a imagem e o núcleo de T sejam idênticos. Demonstrar que n é par. (Dar um exemplo de uma tal transformação linear.)

13. Seja V um espaço vetorial e T uma transformação linear de V em V. Demonstrar que as duas afirmações seguintes sobre T são equivalentes:

(a) A interseção da imagem de T com o núcleo de T é o subespaço nulo de V.
(b) Se $T(T\alpha) = 0$, então $T\alpha = 0$.

3.2 A Álgebra das Transformações Lineares

No estudo das transformações lineares de V em W, é de importância fundamental o fato de que o conjunto dessas transformações herda uma estrutura natural de espaço vetorial. O conjunto das transformações lineares de um espaço V em si mesmo possui uma estrutura algébrica mais rica pois a composição usual de funções fornece uma "multiplicação" dessas transformações. Nesta seção exploraremos essas idéias.

Teorema 4. *Sejam* V *e* W *espaços vetoriais sobre o corpo* F. *Sejam* T *e* U *transformações lineares de* V *em* W. *A função* (T + U) *definida por*

$$(T + U)(\alpha) = T\alpha + U\alpha$$

é uma transformação linear de V *em* W. *Se* c *é um elemento qualquer de* F, *a função* (cT) *definida por*

$$(cT)(\alpha) = c(T\alpha)$$

é uma transformação linear de V *em* W. *O conjunto das transformações lineares de* V *em* W, *munido da adição e multiplicação escalar acima definida, é um espaço vetorial sobre o corpo* F.

Demonstração. Suponhamos que T e U sejam transformações lineares de V em W e definamos $(T+U)$ como acima. Então

$$\begin{aligned}(T + U)(c\alpha + \beta) &= T(c\alpha + \beta) + U(c\alpha + \beta)\\ &= c(T\alpha) + T\beta + c(U\alpha) + U\beta\\ &= c(T\alpha + U\alpha) + (T\beta + U\beta)\\ &= c(T + U)(\alpha) + (T + U)(\beta)\end{aligned}$$

o que mostra que $(T + U)$ é uma transformação linear. Analogamente,

$$\begin{aligned}(cT)(d\alpha + \beta) &= c[T(d\alpha + \beta)]\\ &= c[d(T\alpha) + T\beta]\\ &= cd(T\alpha) + c(T\beta)\\ &= d[c(T\alpha)] + c(T\beta)\end{aligned}$$

mostrando que (cT) é uma transformação linear.

Para verificar que o conjunto das transformações lineares de V em W (munido destas duas operações) é um espaço vetorial, é necessário verificar diretamente cada uma das condições sobre a adição de vetores e a multiplicação escalar. Deixamos a parte principal disto a cargo do leitor e contentamo-nos com este comentário: O vetor nulo deste espaço será a transformação nula, que leva todo vetor de V no vetor nulo de W; cada uma das propriedades das duas operações decorre diretamente da propriedade correspondente das operações no espaço W.

Talvez devamos mencionar outra maneira de considerar este teorema. Se se define soma e múltiplo escalar como fizemos acima, então o conjunto de *todas* as funções de V em W torna-se um espaço vetorial sobre o corpo F. Isto nada tem a ver com o fato de V ser um espaço vetorial, mas apenas com o fato de V ser um conjunto não-vazio. Quando V é um espaço vetorial podemos definir uma transformação linear de V em W e o Teorema 4 diz que as transformações lineares formam um subespaço do espaço de todas as funções de V em W.

Indicaremos o espaço das transformações lineares de V em W por $L(V, W)$. Lembramos novamente ao leitor que $L(V, W)$ está definido somente para V e W espaços vetoriais sobre o mesmo corpo.

Teorema 5. *Seja* V *um espaço vetorial* n*-dimensional sobre o corpo* F *e seja* W *um espaço vetorial* m*-dimensional sobre* F. *Então o espaço* L(V, W) *é de dimensão finita e tem dimensão* mn.

Demonstração. Sejam

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{e} \quad \mathscr{B}' = \{\beta_1, \ldots, \beta_m\}$$

bases ordenadas de V e W, respectivamente. Para cada par de inteiros (p, q) com $1 \leq p \leq m$ e $1 \leq q \leq n$, definamos uma transformação linear $E^{p,q}$ de V em W por

$$\begin{aligned} E^{p,q}(\alpha_i) &= \begin{cases} 0, & \text{se} \quad i \neq q \\ \beta_p, & \text{se} \quad i = q \end{cases} \\ &= \delta_{iq}\beta_p. \end{aligned}$$

De acordo com o Teorema 1, existe uma única transformação linear de V em W que satisfaz estas condições. Afirmamos que as mn transformações $E^{p,q}$ formam uma base de $L(V, W)$.

Seja T uma transformação linear de V em W. Para cada j, $1 \leq j \leq n$, sejam $A_{1j}, \ldots, A_{mj}$ as coordenadas do vetor $T\alpha_j$ em relação à base ordenada $\mathscr{B}'$, isto é,

$$T\alpha_j = \sum_{p=1}^{m} A_{pj}\beta_p. \tag{3-1}$$

Desejamos mostrar que

$$T = \sum_{p=1}^{m} \sum_{q=1}^{n} A_{pq}E^{p,q} \tag{3-2}$$

Seja U a transformação linear no segundo membro de (3-2). Então para cada j

$$\begin{aligned} U\alpha j &= \sum_p \sum_q A_{pq}E^{p,q}(\alpha_i) \\ &= \sum_p \sum_q A_{pq}\delta_{jq}\beta_p \\ &= \sum_{p=1}^{m} A_{pj}\beta_p \\ &= T\alpha_j \end{aligned}$$

e, conseqüentemente, $U = T$. Agora (3-2) mostra que as $E^{p,q}$ geram $L(V, W)$; precisamos demonstrar que elas são independentes. Mas isto é evidente pelo que fizemos acima, pois, se a transformação

$$U = \sum_p \sum_q A_{pq}E^{p,q}$$

é a transformação nula, então $U\alpha_j = 0$ para cada j, portanto

$$\sum_{p=1}^{m} A_{pj}\beta_p = 0$$

e a independência dos β_p implica que $A_{pj} = 0$ para todos p e j.

Teorema 6. *Sejam* V, W *e* Z *espaços vetoriais sobre o corpo* F. *Seja* T *uma transformação linear de* V *em* W *e* U *uma transformação*

linear de W *em* Z. *Então, a função composta* UT, *definida por* (UT) (α) = U (T(α)) *é uma transformação linear de* V *em* Z.

Demonstração

$$\begin{aligned}(UT)(c\alpha + \beta) &= U[T(c\alpha + \beta)] \\ &= U(cT\alpha + T\beta) \\ &= c[U(T\alpha)] + U(T\beta) \\ &= c(UT)(\alpha) + (UT)(\beta).\end{aligned}$$

No que segue, estaremos primordialmente preocupados com transformações lineares de um espaço vetorial em si mesmo. Como teríamos de escrever a todo instante "T é uma transformação linear de V em V", substituiremos isto por "T é um operador linear sobre V".

Definição. *Se* V *é um espaço vetorial sobre o corpo* F, *um* **operador linear sobre** V *é uma transformação linear de* V *em* V.

Aplicando o Teorema 6 para $V = W = Z$, de modo que U e T sejam operadores lineares sobre o espaço V, vemos que a composta UT é ainda um operador linear sobre V. Assim, o espaço $L(V, V)$ possui uma "multiplicação", definida sobre si por meio da composição. Neste caso, o operador TU também está definido e devemos notar que em geral $UT \neq TU$, isto é, $UT - TU \neq 0$. Devemos notar particularmente o fato de que se T é um operador linear sobre V, então podemos compor T com T. Usaremos a notação $T^2 = TT$ e, em geral, $T^n = T \ldots T$ (n vezes) para $n = 1, 2, 3, \ldots$. Definimos $T^0 = I$ se $T \neq 0$.

Lema. *Seja* V *um espaço vetorial sobre o corpo* F; *sejam* U, T_1 *e* T_2 *operadores lineares sobre* V; *seja* c *um elemento de* F.

(a) $IU = UI = U$;

(b) $U(T_1 + T_2) = UT_1 + UT_2$; $(T_1 + T_2)U = T_1U + T_2U$;

(c) $c(UT_1) = (cU)T_1 = U(cT_1)$.

Demonstração. (a) Esta propriedade da função idêntica é óbvia. Enunciamo-la aqui por mera questão de ênfase.

(b) $$\begin{aligned}[U(T_1 + T_2)](\alpha) &= U[(T_1 + T_2)(\alpha)] \\ &= U(T_1\alpha + T_2\alpha)\end{aligned}$$

$$\begin{aligned} &= U(T_1\alpha) + U(T_2\alpha) \\ &= (UT_1)(\alpha) + (UT_2)(\alpha) \end{aligned}$$

de modo que $U(T_1 + T_2) = UT_1 + UT_2$. Além disso

$$\begin{aligned} [(T_1 + T_2)U](\alpha) &= (T_1 + T_2)(U\alpha) \\ &= T_1(U\alpha) + T_2(U\alpha) \\ &= (T_1U)(\alpha) + (T_2U)(\alpha) \end{aligned}$$

de modo que $(T_1 + T_2)U = T_1U + T_2U$. (O leitor pode notar que para as demonstrações destas duas leis distributivas não foi usado o fato de T_1 e T_2 serem lineares, e que para a demonstração da segunda não foi usado tampouco o fato de U ser linear.)

(c) Deixamos a demonstração da parte (c) a cargo do leitor.

Este lema e uma parte do Teorema 5 nos dizem que o espaço vetorial $L(V, V)$, munido da operação de composição, é o que se conhece por uma álgebra linear com elemento unidade. Discutiremos isto no Capítulo 4.

Exemplo 8. Se A é uma $m \times n$ matriz com elementos em F, temos a transformação linear T definida por $T(X) = AX$, de $F^{n \times 1}$ em $F^{m \times 1}$. Se B é uma $p \times m$ matriz, temos a transformação linear U de $F^{m \times 1}$ em $F^{p \times 1}$ definida por $U(Y) = BY$. É fácil descrever a composta UT:

$$\begin{aligned} (UT)(X) &= U(T(X)) \\ &= U(AX) \\ &= B(AX) \\ &= (BA)X. \end{aligned}$$

Assim, UT é "a multiplicação à esquerda pela matriz produto BA".

Exemplo 9. Seja F um corpo e V o espaço vetorial das funções polinomiais de F em F. Seja D o operador derivação definido no Exemplo 2 e T o operador linear "multiplicação por x":

$$(Tf)(x) = xf(x).$$

Então $DT \neq TD$. Na verdade, o leitor deverá achar fácil verificar que $DT - TD = I$, o operador idêntico.

Mesmo não sendo comutativa, a "multiplicação" que temos sobre $L(V, V)$ está bastante relacionada com as operações de espaço vetorial de $L(V, V)$.

Exemplo 10. Seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base ordenada do espaço vetorial V. Consideremos os operadores lineares $E^{p,q}$ que apareceram na demonstração do Teorema 5:

$$E^{p,q}(\alpha_i) = \delta_{iq}\alpha_p.$$

Estes n^2 operadores lineares formam uma base do espaço dos operadores lineares sobre V. O que vem a ser $E^{p,q}E^{r,s}$? Temos

$$\begin{aligned}(E^{p,q}E^{r,s})(\alpha_i) &= E^{p,q}(\delta_{is}\alpha_r)\\ &= \delta_{is}E^{p,q}(\alpha_r)\\ &= \delta_{is}\delta_{rq}\alpha_p.\end{aligned}$$

Portanto,

$$E^{p,q}E^{r,s} = \begin{cases} 0, & \text{se} \quad r \neq q \\ E^{p,s}, & \text{se} \quad q = r.\end{cases}$$

Seja T um operador linear sobre V. Vimos na demonstração do Teorema 5 que se

$$\begin{aligned} A_j &= [T\alpha_j]_{\mathscr{B}} \\ A &= [A_1, \ldots, A_n]\end{aligned}$$

então

$$T = \sum_p \sum_q A_{pq}E^{p,q}.$$

Se

$$U = \sum_r \sum_s B_{rs}E^{r,s}$$

for um outro operador linear sobre V então pelo último lema

$$\begin{aligned} TU &= \Big(\sum_p \sum_q A_{pq}E^{p,q}\Big)\Big(\sum_r \sum_s B_{rs}E^{r,s}\Big) \\ &= \sum_p \sum_q \sum_r \sum_s A_{pq}B_{rs}E^{p,q}E^{r,s}.\end{aligned}$$

Como observamos, os únicos termos que permanecem nesta soma enorme são os termos nos quais $q = r$ e como $E^{p,r}E^{r,s} = E^{p,s}$, temos

$$TU = \sum_p \sum_s (\sum_r A_{pr}B_{rs}) E^{p,s}$$

$$= \sum_p \sum_s (AB)_{ps} E^{p,s}.$$

Assim, o efeito de compor T e U é o de multiplicar as matrizes A e B.

Na nossa discussão sobre operações algébricas com transformações lineares, nada dissemos sobre inversibilidade. Uma questão específica de interesse é a seguinte: para quais operadores lineares T sobre o espaço V existe um operador linear T^{-1} tal que $TT^{-1} = T^{-1}T = I$?

A função T de V em W se diz **inversível** se existir uma função U de W em V tal que UT é a função idêntica sobre V e TU é a função idêntica sobre W. Se T for inversível, a função U é única e é representada por T^{-1}. (Veja Apêndice.) Além do mais, T é inversível se, e somente se,

1. T é injetora, isto é, $T\alpha = T\beta$ implica $\alpha = \beta$;
2. T é sobrejetora, isto é, a imagem de T é W.

Teorema 7. *Sejam* V *e* W *espaços vetoriais sobre o corpo* F *e seja* T *uma transformação linear de* V *em* W. *Se* T *é inversível então a função inversa* T^{-1} *é uma transformação linear de* W *sobre* V.

Demonstração. Nós nos repetiremos aqui por questão de ênfase. Quando T é injetora e sobrejetora, existe uma função inversa T^{-1}, determinada de modo único, que leva W sobre V tal que $T^{-1}T$ é a função idêntica de V e TT^{-1} é a função idêntica de W. O que estamos demonstrando aqui é que, se uma função linear T é inversível, então a inversa T^{-1} também é linear.

Sejam β_1 e β_2 vetores em W e seja c um escalar. Queremos mostrar que

$$T^{-1}(c\beta_1 + \beta_2) = cT^{-1}\beta_1 + T^{-1}\beta_2.$$

Seja $\alpha_i = T^{-1}\beta_i$, $i = 1, 2$, isto é, seja α_i o único vetor em V tal que $T\alpha_i = \beta_i$. Como T é linear,

$$\begin{aligned} T(c\alpha_1 + \alpha_2) &= cT\alpha_1 + T\alpha_2 \\ &= c\beta_1 + \beta_2. \end{aligned}$$

Assim, $c\alpha_1 + \alpha_2$ é o único vetor em V que é levado por T em $c\beta_1 + \beta_2$, portanto

$$\begin{aligned} T^{-1}(c\beta_1 + \beta_2) &= c\alpha_1 + \alpha_2 \\ &= c(T^{-1}\beta_1) + T^{-1}\beta_2 \end{aligned}$$

e T^{-1} é linear.

Suponhamos que T seja uma transformação linear inversível de V sobre W e que U seja uma transformação linear inversível de W sobre Z. Então UT é inversível e $(UT)^{-1} = T^{-1}U^{-1}$. Esta conclusão não requer a linearidade, nem a verificação em separado de que UT seja injetora e sobrejetora. Ela somente requer a verificação de que $T^{-1}U^{-1}$ seja ao mesmo tempo uma inversa à esquerda e à direita de UT.

Se T for linear, então $T(\alpha - \beta) = T\alpha - T\beta$; logo, $T\alpha = T\beta$ se, e somente e, $T(\alpha - \beta) = 0$. Isto simplifica imensamente a verificação de que T é injetora. Diremos que uma transformação linear T é **não-singular** se $T\gamma = 0$ implicar $\gamma = 0$, isto é, se o núcleo de T for $\{0\}$. Evidentemente, T é injetora se, e somente se, T for não singular. A importância desta observação é que transformações lineares não-singulares são aquelas que preservam a independência linear.

Teorema 8. *Seja* T *uma transformação linear de* V *em* W. *Então* T *é não-singular se, e somente se,* T *leva todo subconjunto linearmente independente de* V *sobre um subconjunto linearmente independente de* W.

Demonstração. Suponhamos primeiro que T seja não-singular. Seja S um subconjunto linearmente independente de V. Se $\alpha_1, \ldots, \alpha_k$ são vetores em S, então os vetores $T\alpha_1, \ldots, T\alpha_k$ são linearmente independentes, pois se

$$c_1(T\alpha_1) + \ldots + c_k(T\alpha_k) = 0$$

então

$$T(c_1\alpha_1 + \ldots + c_k\alpha_k) = 0$$

e como T é não-singular

$$c_1\alpha_1 + \ldots + c_k\alpha_k = 0$$

do que segue que cada $c_i = 0$ pois S é um conjunto independente. Este argumento mostra que a imagem de S por meio de T é independente.

Suponhamos que T leve subconjuntos independentes sobre subconjuntos independentes. Seja α um vetor não-nulo em V. Então o conjunto S constituído apenas pelo vetor α é independente. A imagem de S é o conjunto constituído apenas pelo vetor $T\alpha$ e este conjunto é independente. Portanto $T\alpha \neq 0$, pois o conjunto constituído apenas pelo vetor nulo é dependente. Isto mostra que o núcleo de T é o subespaço nulo, isto é, T é não-singular.

Exemplo 11. Seja F um subcorpo do corpo dos números complexos (ou um corpo de característica zero) e seja V o espaço das funções polinomiais sobre F. Consideremos o operador derivação D e o operador T, "multiplicação por x", do Exemplo 9. Como D leva todas as constantes em zero, D é singular; no entanto, V não é de dimensão finita, a imagem de D é todo o espaço V e é possível definir um inverso à direita para D. Por exemplo, se E for o operador integração indefinida:

$$E(c_0 + c_1x + \dots + c_nx^n) = c_0x + \frac{1}{2}c_1x^2 + \dots + \frac{1}{n+1}c_nx^{n+1}$$

então E é um operador linear sobre V e $DE = I$. Por outro lado, $ED \neq I$ pois ED leva as constantes em zero. O operador T está, no que podemos chamar de situação reversa. Se $xf(x) = 0$ para todo x, então $f = 0$. Portanto, T é não-singular e é possível determinar um inverso à esquerda para T. Por exemplo, se U for a operação "remover o termo constante e dividir por x":

$$U(c_0 + c_1x + \dots + c_nx^n) = c_1 + c_2x + \dots + c_nx^{n-1}$$

então U é um operador linear sobre V e $UT = I$. Mas $TU \neq I$ pois toda a função na imagem de TU está na imagem de T que é o espaço das funções polinomiais f tais que $f(0) = 0$.

Exemplo 12. Seja F um corpo e seja T o operador linear sobre F^2 definido por

$$T(x_1, x_2) = (x_1 + x_2, x_1).$$

Então T é não singular, pois se $T(x_1, x_2) = 0$, temos

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_1 &= 0 \end{aligned}$$

de modo que $x_1 = x_2 = 0$. Vejamos também que T é sobrejetor; de fato, seja (z_1, z_2) um vetor arbitrário em F^2. Para mostrar que (z_1, z_2) está na imagem de T precisamos determinar escalares x_1 e x_2 tais que

$$\begin{aligned} x_1 + x_2 &= z_1 \\ x_1 &= z_2 \end{aligned}$$

e a solução é evidentemente $x_1 = z_2$, $x_2 = z_1 - z_2$. Este último cálculo nos dá uma fórmula explícita para T^{-1}, a saber,

$$T^{-1}(z_1, z_2) = (z_2, z_1 - z_2).$$

Vimos no Exemplo 11 que uma transformação linear pode ser não-singular sem ser sobrejetora e pode ser sobrejetora sem ser não-singular. O exemplo acima ilustra um caso importante onde isto não pode acontecer.

Teorema 9. *Sejam* V *e* W *espaços vetoriais de dimensão finita sobre o corpo* F *tais que* dim V = *dim* W. *Se* T *é uma transformação linear de* V *em* W, *as seguintes afirmações são equivalentes:*

(i) T *é inversível.*

(ii) T *é não-singular.*

(iii) T é *sobrejetora, isto é, a imagem de* T *é* W.

Demonstração. Seja $n = \dim V = \dim W$. Pelo Teorema 2 sabemos que

$$\text{posto } (T) + \text{nulidade } (T) = n.$$

Ora, T é não singular se, e somente se, nulidade $(T) = 0$, e (como $n = \dim W$) a imagem de T é W se, e somente se, posto $(T) = n$. Como o posto mais a nulidade é n, a nulidade é 0 precisamente quando o posto for n. Portanto T é não-singular se, e somente se, $T(V) = W$. Assim, se uma das condições (ii) ou (iii) se verificar, a outra também se verificará e T será inversível.

O leitor deve tomar cuidado de não aplicar o Teorema 9 salvo na presença de dimensionalidade finita e com $\dim V = \dim W$.

Com as hipóteses do Teorema 9, as condições (i), (ii) e (iii) são também equivalentes às seguintes:

(iv) *Se* $\{\alpha_1, \ldots, \alpha_n\}$ *é uma base arbitrária de* V, *então* $\{T\alpha_1, \ldots, T\alpha_n\}$ *é uma base de* W.

(v) *Existe pelo menos uma base* $\{\alpha_1, \ldots, \alpha_n\}$ *de* V *tal que* $\{T\alpha_1, \ldots, T\alpha_n\}$ *seja uma base de* W.

Daremos uma demonstração da equivalência das cinco condições, que contém uma demonstração diferente, da equivalência de (i), (ii) e (iii).

(i)→(ii). Se T é inversível, T é não-singular. (ii)→(iii). Suponhamos que T seja não-singular. Seja $\{\alpha_1, \ldots, \alpha_n\}$ uma base de V. Pelo Teorema 8, $\{T\alpha_1, \ldots, T\alpha_n\}$ é um conjunto linearmente independente de vetores em W e como a dimensão de W também é n, este conjunto de vetores é uma base de W. Agora seja β um vetor arbitrário em W. Existem escalares $c_1, \ldots, c_n$ tais que

$$\begin{aligned}\beta &= c_1(T\alpha_1) + \ldots + c_n(T\alpha_n)\\ &= T(c_1\alpha_1 + \ldots + c_n\alpha_n)\end{aligned}$$

o que mostra que β está na imagem de T. (iii)→(iv). Suponhamos agora que T seja sobrejetora. Se $\{\alpha_1, \ldots, \alpha_n\}$ é uma base arbitrária de V, os vetores $T\alpha_1, \ldots, T\alpha_n$ geram a imagem de T, que é todo o espaço W por hipótese. Como a dimensão de W é n, estes n vetores precisam ser linearmente independentes, isto é, precisam formar uma base de W. (iv)→(v). Não requer nenhum comentário. (v)→(i). Suponhamos que exista alguma base $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que $\{T\alpha_1, \ldots, T\alpha_n\}$ seja uma base de W. Como os $T\alpha_i$ geram W, é evidente que a imagem de T coincide com W. Se $\alpha = c_1\alpha_1 + \ldots + c_n\alpha_n$ está no núcleo de T, então

$$T(c_1\alpha_1 + \ldots + c_n\alpha_n) = 0$$

ou

$$c_1(T\alpha_1) + \ldots + c_n(T\alpha_n) = 0$$

e como os $T\alpha_i$ são independentes, cada $c_i = 0$ e assim $\alpha = 0$. Mostramos que a imagem de T é W e que T é não-singular, logo T é inversível.

O conjunto dos operadores lineares inversíveis sobre um espaço V, com a operação de composição, fornece um belo exemplo do que é conhecido em álgebra por um "grupo". Apesar de que

não teremos tempo para discutir grupos com quaisquer pormenores, daremos pelo menos a definição.

Definição. *Um* **grupo** *consiste do seguinte:*

(1) *Um conjunto* G;

(2) *Uma regra (ou operação) que associa a cada par de elementos* x, y *em* G *um elemento* xy *em* G *de uma maneira tal que*

(a) x(yz) = (xy)z, *para todos* x, y *e* z *em* G *(associatividade)*;

(b) *existe um elemento* e *em* G *tal que* ex = xe = x, *para todo* x *em* G;

(c) *a cada elemento* x *em* G *corresponde um elemento* x^{-1} *em* G *tal que* $xx^{-1} = x^{-1}x = e$.

Vimos que a composição $(U, T) \rightarrow UT$ associa a cada par de operadores lineares inversíveis sobre um espaço V outro operador inversível sobre V. A composição é uma operação associativa. O operador idêntico I satisfaz $IT = TI = T$ para todo T e para um T inversível existe (pelo Teorema 7) um operador linear inversível T^{-1} tal que $TT^{-1} = T^{-1}T = I$. Portanto o conjunto dos operadores lineares inversíveis sobre V, munido desta operação, é um grupo. O conjunto das $n \times n$ matrizes inversíveis com a multiplicação de matrizes como a operação é outro exemplo de um grupo. Um grupo é dito **comutativo** se satisfaz a condição $xy = yx$ para todos x e y. Os dois exemplos que demos acima não são, em geral, grupos comutativos. Freqüentemente indica-se a operação num grupo comutativo por $(x, y) \rightarrow x + y$ em lugar de $(x, y) \rightarrow xy$ e usa-se então o símbolo 0 para o elemento "unidade" e. O conjunto dos vetores de um espaço vetorial, com a operação de adição de vetores, é um grupo comutativo. Um corpo pode ser descrito como um conjunto com duas operações, denominadas adição e multiplicação, que é um grupo comutativo em relação à adição e no qual os elementos não-nulos formam um grupo comutativo em relação à multiplicação, valendo a lei distributiva $x(y + z) = xy + xz$.

Exercícios

1. Sejam T e U os operadores lineares sobre R^2 definidos por

$$T(x_1, x_2) = (x_2, x_1) \quad \text{e} \quad U(x_1, x_2) = (x_1, 0).$$

(a) Como você descreveria T e U geometricamente?

(b) Dar regras como as que definem T e U para cada uma das transformações $(U + T)$, UT, TU, T^2 U^2.

2. Seja T o (único) operador linear sobre C^3 para o qual

$$T\varepsilon_1 = (1, 0, i), \quad T\varepsilon_2 = (0, 1, 1), \quad T\varepsilon_3 = (i, 1, 0).$$

T é inversível?

3. Seja T o operador linear sobre R^3 definido por

$$T(x_1, x_2, x_3) = (3x_1, x_1 - x_2, 2x_1 + x_2 + x_3).$$

T é inversível? Em caso afirmativo, determinar uma regra para T^{-1} como a que define T.

4. Para o operador linear T do Exercício 3, demonstrar que

$$(T^2 - I)(T - 3I) = 0.$$

5. Seja $C^{2\times 2}$ o espaço vetorial complexo das 2×2 matrizes como elementos complexos. Seja

$$B = \begin{bmatrix} 1 & -1 \\ -4 & 4 \end{bmatrix}$$

e seja T o operador linear sobre $C^{2\times 2}$ definido por $T(A) = BA$. Qual é o posto de T? Descrever T^2.

6. Seja T uma transformação linear de R^3 em R^2 e seja U uma transformação linear de R^2 em R^3. Demonstrar que a transformação linear UT não é inversível. Generalizar o teorema.

7. Determinar dois operadores lineares T e U sobre R^2 tais que $TU = 0$ mas $UT \neq 0$.

8. Seja V um espaço vetorial sobre o corpo F e T um operador linear sobre V. Se $T^2 = 0$, o que se pode dizer sobre a relação entre a imagem de T e o núcleo de T? Dar um exemplo de um operador linear T sobre R^2 tal que $T^2 = 0$ mas $T \neq 0$.

9. Seja T um operador linear sobre o espaço vetorial V de dimensão finita. Suponhamos que exista um operador linear U sobre V tal que $TU = I$. Demonstrar que T é inversível e $U = T^{-1}$. Dar um exemplo que mostre que isto é falso quando V não é de dimensão finita. (*Sugestão*: Seja $T = D$, o operador derivação sobre o espaço das funções polinomiais.)

10. Seja A uma $m \times n$ matriz com elementos em F e seja T a transformação linear de $F^{n\times 1}$ em $F^{m\times 1}$ definida por $T(X) = AX$. Mostrar que se $m < n$ pode acontecer que T seja sobrejetora sem ser não-singular. Analogamente, mostrar que se $m > n$ podemos ter T não-singular mas não sobrejetora.

11. Seja V um espaço vetorial de dimensão finita e seja T um operador linear sobre V. Suponhamos que posto $(T^2) =$ posto (T). Demonstrar que a imagem e o núcleo de T são disjuntos, isto é, possuem em comum apenas o vetor nulo.

12. Sejam p, m e n inteiros positivos e F um corpo. Seja V o espaço das $m \times n$ matrizes sobre F e W o espaço das $p \times n$ matrizes sobre F. Seja B uma $p \times m$ matriz fixa e seja T a transformação linear de V em W definida por $T(A) = BA$. Demonstrar que T é inversível se, e somente se, $p = m$ e B é uma $m \times m$ matriz inversível.

3.3. Isomorfismo

Se V e W são espaços vetoriais sobre o corpo F, uma transformação linear bijetora (injetora e sobrejetora) T de V em W é denominada um **isomorfismo de** V **em** W. Se existir um isomorfismo de V em W, diremos que V é **isomorfo a** W.

Notemos que V é trivialmente isomorfo a V, pois o operador idêntico é um isomorfismo de V em V. Além disso, se V é isomorfo a W por meio de um isomorfismo T, então W é isomorfo a V uma vez que T^{-1} é um isomorfismo de W em V. O leitor deverá achar fácil verificar que se V é isomorfo a W e W é isomorfo a Z, então V é isomorfo a Z. Em suma, o isomorfismo é uma relação de equivalência sobre a classe dos espaços vetoriais. Se existir um isomorfismo de V em W, poderemos às vezes dizer que V e W são isomorfos, em vez de dizer que V é isomorfo a W. Isto não causará confusão alguma porque V é isomorfo a W se, e somente se, W é isomorfo a V.

Teorema 10. *Todo espaço vetorial* n*-dimensional sobre o corpo* F *é isomorfo ao espaço* F^n.

Demonstração. Seja V um espaço n-dimensional sobre o corpo F e seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base ordenada de V. Definamos uma função T de V em F^n, como segue: Se α está em V, seja $T\alpha$ a n-upla $(x_1, \ldots, x_n)$ das coordenadas de α em relação à base ordenada $\mathscr{B}$, isto é, a n-upla tal que

$$\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n.$$

Em nossa discussão de coordenadas no Capítulo 2, verificamos que esta T é linear, injetora e leva V sobre F^n.

Para muitos objetivos freqüentemente consideram-se espaços vetoriais isomorfos como sendo "o mesmo", apesar de que os vetores e as operações nos espaços possam ser bem diferentes, isto é, freqüentemente identificamos espaços isomorfos. Não tentaremos fazer uma longa discussão sobre esta idéia no momento mas deixaremos a compreensão do isomorfismo e do sentido no qual espaços isomorfos são "o mesmo" crescerem à medida que continuemos nosso estudo de espaços vetoriais.

Faremos alguns comentários breves. Suponhamos que T seja um isomorfismo de V em W. Se S é um subconjunto de V, o Teo-

rema 8 nos diz que S é linearmente independente se, e somente se, o conjunto $T(S)$ em W é independente. Portanto, ao decidirmos se S é independente não importa se consideramos S ou $T(S)$. A partir disto vê-se que um isomorfismo "conserva a dimensão", isto é, todo subespaço de V de dimensão finita tem a mesma dimensão que sua imagem por meio de T. Eis uma ilustração muito simples dessa idéia. Suponhamos que A seja uma $m \times n$ matriz sobre o corpo F. Na verdade demos duas definições do espaço-solução da matriz A. O primeiro é o conjunto das n-uplas $(x_1, \ldots, x_n)$ em F^n que satisfazem cada uma das equações do sistema $AX = 0$. O segundo é o conjunto das $n \times 1$ matrizes colunas X tais que $AX = 0$. O primeiro espaço-solução é, portanto, um subespaço de F^n e o segundo é um subespaço do espaço de todas as $n \times 1$ matrizes sobre F. Agora existe um isomorfismo evidente entre F^n e $F^{n \times 1}$, a saber,

$$(x_1, \ldots, x_n) \to \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

Por meio deste isomorfismo, o primeiro espaço-solução de A é levado sobre o segundo espaço-solução. Estes espaços têm a mesma dimensão, portanto se quisermos demonstrar um teorema sobre a dimensão do espaço-solução, não importará qual espaço resolvamos discutir. Na verdade, o leitor provavelmente não objetaria se resolvêssemos identificar F^n com o espaço das $n \times 1$ matrizes. Poderemos fazê-lo quando for conveniente, e quando não o for não o faremos.

Exercícios

1. Seja V o conjunto dos números complexos e seja F o corpo dos números reais. Com as operações usuais, V é um espaço vetorial sobre F. Descrever explicitamente um isomorfismo deste espaço em R^2.

2. Seja V um espaço vetorial sobre o corpo dos números complexos e suponhamos que exista um isomorfismo T de V em C^3. Sejam α_1, α_2, α_3, α_4 vetores em V tais que

$$T\alpha_1 = (1, 0, i) \qquad T\alpha_2 = (-2, 1+i, 0),$$
$$T\alpha_3 = (-1, 1, 1), \qquad T\alpha_4 = (\sqrt{2}, i, 3).$$

(a) α_1 está no subespaço gerado por α_2 e α_3?
(b) Seja W_1 o subespaço gerado por α_1 e α_2 e seja W_2 o subespaço gerado por α_3 e α_4. Qual é a interseção de W_1 com W_2?
(c) Determinar uma base do subespaço de V gerado pelos quatro vetores α_j.

3. Seja W o conjunto das 2×2 matrizes hermitianas, complexas, isto é, o conjunto das 2×2 matrizes complexas A tais que $A_{ij} = \overline{A_{ji}}$ (a barra indica conjugação complexa). Como destacamos no Exemplo 6 do Capítulo 2, W é um espaço vetorial sobre o corpo dos números *reais*, em relação às operações usuais. Verificar que

$$(x, y, z, t) \to \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

é um isomorfismo de R^4 em W.

4. Mostrar que $F^{m \times n}$ é isomorfo a F^{mn}.

5. Seja V o conjunto dos números complexos considerado como um espaço vetorial sobre o corpo dos números reais (Exercício 1). Definamos uma função T de V no espaço das 2×2 matrizes reais, como segue. Se $z = x + iy$ com x e y números reais, então

$$T(z) = \begin{bmatrix} x + 7y & 5y \\ -10y & x - 7y \end{bmatrix}.$$

(a) Verificar que T é uma transformação linear (real) injetora de V no espaço das 2×2 matrizes.
(b) Verificar que $T(z_1 z_2) = T(z_1)\, T(z_2)$.
(c) Como você descreveria a imagem de T?

6. Sejam V e W espaços vetoriais de dimensão finita sobre o corpo F. Demonstrar que V e W são isomorfos se, e somente se, $\dim V = \dim W$.

7. Sejam V e W espaços vetoriais sobre o corpo F e seja U um isomorfismo de V em W. Demonstrar que $T \to UTU^{-1}$ é um isomorfismo de $L(V, V)$ em $L(W, W)$.

3.4 Representação de Transformações por Matrizes

Seja V um espaço vetorial n-dimensional sobre o corpo F e seja W um espaço vetorial m-dimensional sobre F. Sejam $\mathscr{B} = \{\alpha_1, \dots, \alpha_n\}$ uma base ordenada de V e $\mathscr{B}' = \{\beta_1, \dots, \beta_m\}$ uma base ordenada de W. Se T é uma transformação linear arbitrária de V em W, então T é determinada por seu efeito sobre os vetores α_j. Cada um dos n vetores $T\alpha_j$ pode ser expresso de modo único como uma combinação linear

$$T\alpha_j = \sum_{i=1}^{m} A_{ij}\beta_i \tag{3-3}$$

dos β_i, sendo os escalares $A_{1j}, \ldots, A_{mj}$ as coordenadas de $T\alpha_j$ em relação à base ordenada $\mathscr{B}'$. Conseqüentemente, a transformação T é determinada pelos mn escalares A_{ij} por meio das fórmulas (3-3). A $m \times n$ matriz A definida por $A(i, j) = A_{ij}$ é denominada **a matriz de T em relação ao par de bases ordenadas $\mathscr{B}$ e $\mathscr{B}'$**. Nosso trabalho imediato será o de compreender explicitamente como a matriz A determina a transformação linear T.

Se $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$ é um vetor em V, então

$$\begin{aligned} T\alpha &= T\left(\sum_{j=1}^{n} x_j\alpha_j\right) \\ &= \sum_{j=1}^{n} x_j(T\alpha_j) \\ &= \sum_{j=1}^{n} c_j \sum_{i=1}^{m} A_{ij}\beta_i \\ &= \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j\right)\beta_i. \end{aligned}$$

Se X é a matriz das coordenadas de α em relação à base ordenada $\mathscr{B}$, o cálculo acima mostra que AX é a matriz das coordenadas do vetor $T\alpha$ em relação à base ordenada $\mathscr{B}'$, uma vez que o escalar

$$\sum_{j=1}^{n} A_{ij}x_j$$

é o elemento da i-ésima linha da matriz coluna AX. Observemos também que se A é uma $m \times n$ matriz arbitrária sobre o corpo F então

$$T\left(\sum_{j=1}^{n} x_j\alpha_j\right) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j\right) {}_j\beta_i \tag{3-4}$$

define uma transformação linear T de V em W, cuja matriz é A, em relação a $\mathscr{B}$, $\mathscr{B}'$. Resumindo formalmente:

Teorema 11. *Seja* V *um espaço vetorial* n*-dimensional sobre o corpo* F *e* W *um espaço vetorial* m*-dimensional sobre* F. *Seja* $\mathscr{B}$ *uma base ordenada de* V *e* $\mathscr{B}'$ *uma base ordenada de* W. *Para cada trans-*

formação linear T *de* V *em* W, *existe uma* m × n *matriz* A *com elementos em* F *tal que*

$$[T\alpha]_{\mathscr{B}'} = A[\alpha]_{\mathscr{B}}$$

para todo vetor α *em* V. *Além disso,* T → A *é uma correspondência bijetora entre o conjunto das transformações lineares de* V *em* W *e o conjunto das* m × n *matrizes sobre o corpo* F.

A matriz A, associada a T no Teorema 11, é denominada **matriz de T em relação às bases ordenadas $\mathscr{B}$ e $\mathscr{B}'$.** Notemos que pela Equação (3-3), A é a matriz cujas colunas são dadas por

$$A_j = [T\alpha_j]_{\mathscr{B}'}, \quad j = 1, \ldots, n.$$

Se U for uma outra transformação de V em W e $B = [B_1, \ldots, B_n]$ for a matriz de U em relação às bases ordenadas $\mathscr{B}$ e $\mathscr{B}'$, então $cA + B$ será a matriz de $cT + U$ em relação a $\mathscr{B}$ e $\mathscr{B}'$. Isto é evidente, pois

$$\begin{aligned} cA_j + B_j &= c[T\alpha_j]_{\mathscr{B}'} + [U\alpha_j]_{\mathscr{B}'} \\ &= [cT\alpha_j + U\alpha_j]_{\mathscr{B}'} \\ &= [(cT + U)\alpha_j]_{\mathscr{B}'}. \end{aligned}$$

Teorema 12. *Seja* V *um espaço vetorial* n*-dimensional sobre o corpo* F *e seja* W *um espaço vetorial* m*-dimensional sobre* F. *Para cada par de bases ordenadas* $\mathscr{B}$, $\mathscr{B}'$ *de* V *e* W, *respectivamente, a função que associa a uma transformação linear* T *sua matriz em relação a* $\mathscr{B}$, $\mathscr{B}'$ *é um isomorfismo entre o espaço* L(V, W) *e o espaço das* m × n *matrizes sobre o corpo* F.

Demonstração. Observamos acima que a função em questão é linear e, como está enunciado no Teorema 11, esta função é injetora e leva $L(V, W)$ sobre o conjunto das $m \times n$ matrizes.

Estaremos particularmente interessados na representação por matrizes de transformações lineares de um espaço em si mesmo, isto é, operadores lineares sobre um espaço V. Neste caso, é mais conveniente usar a mesma base ordenada em cada caso, isto é, tomar $\mathscr{B} = \mathscr{B}'$. A matriz representante será então denominada simplesmente a **matriz de T em relação à base ordenada $\mathscr{B}$.** Como este conceito será muito importante para nós, recordaremos sua definição. Se T é um operador linear sobre o espaço vetorial V de dimensão finita e $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ é uma base ordenada de V,

a matriz de T em relação a $\mathscr{B}$ é a $n \times n$ matriz A cujos elementos A_{ij} são definidos pelas equações

$$(3\text{-}5) \qquad T\alpha_j = \sum_{i=1}^{n} A_{ij}\alpha_i, \quad j = 1, \ldots, n.$$

Deve-se ter sempre em mente que esta matriz que representa T depende da base ordenada $\mathscr{B}$ e que existe uma matriz que representa T em relação a cada base ordenada de V. (Para transformações de um espaço em outro a matriz depende de duas bases ordenadas, uma de V e uma de W.) Para não esquecermos esta dependência, usaremos a notação

$$[T]_{\mathscr{B}}$$

para a matriz do operador linear T em relação à base ordenada $\mathscr{B}$. A maneira como esta matriz e a base ordenada descrevem T é que, para cada α em V,

$$[T\alpha]_{\mathscr{B}} = [T]_{\mathscr{B}}[\alpha]_{\mathscr{B}}.$$

Exemplo 13. Seja V o espaço das $n \times 1$ matrizes-colunas sobre o corpo F; seja W o espaço das $m \times 1$ matrizes sobre F; seja A uma $m \times n$ matriz sobre F, fixa. Seja T a transformação linear de V em W definida por $T(X) = AX$. Seja $\mathscr{B}$ a base ordenada de V análoga à base canônica em F^n, isto é, o i-ésimo vetor em $\mathscr{B}$ é a $n \times 1$ matriz X_i com 1 na linha i e com todos os outros elementos nulos. Seja $\mathscr{B}'$ a correspondente base ordenada de W, isto é, o j-ésimo vetor em $\mathscr{B}'$ é a $m \times 1$ matriz Y_j com 1 na linha j e com todos os outros elementos nulos. Então a matriz de T em relação ao par $\mathscr{B}$, $\mathscr{B}'$ é a própria matriz A. Isto é evidente, pois a matriz AX_j é a j-ésima coluna de A.

Exemplo 14. Seja F um corpo e seja T o operador sobre F^2 definido por

$$T(x_1, x_2) = (x_1, 0).$$

É fácil ver que T é um operador linear sobre F^2. Seja $\mathscr{B}$ a base ordenada canônica de F^2, $\mathscr{B} = \{\varepsilon_1, \varepsilon_2\}$. Ora,

$$\begin{aligned} T\varepsilon_1 &= T(1, 0) = (1, 0) = 1\varepsilon_1 + 0\varepsilon_2 \\ T\varepsilon_2 &= T(0, 1) = (0, 0) = 0\varepsilon_1 + 0\varepsilon_2 \end{aligned}$$

de modo que a matriz de T em relação à base ordenada $\mathscr{B}$ é

$$[T]_{\mathscr{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Exemplo 15. Seja V o espaço das funções polinomiais de R em R da forma

$$f(x) = c_0 + c_1x + c_2x^2 + c_3x^3$$

isto é, o espaço das funções polinomiais de grau menor ou igual a 3. O operador derivação D do Exemplo 2 leva V em V, pois D diminui o grau. Seja $\mathscr{B}$ a base ordenada de V formada pelas quatro funções f_1, f_2, f_3, f_4 definidas por $f_j(x) = x^{j-1}$. Então

$$\begin{aligned}
(Df_1)(x) &= 0, & Df_1 &= 0f_1 + 0f_2 + 0f_3 + 0f_4 \\
(Df_2)(x) &= 1, & Df_2 &= 1f_1 + 0f_2 + 0f_3 + 0f_4 \\
(Df_3)(x) &= 2x, & Df_3 &= 0f_1 + 2f_2 + 0f_3 + 0f_4 \\
(Df_4)(x) &= 3x^2, & Df_4 &= 0f_1 + 0f_2 + 3f_3 + 0f_4
\end{aligned}$$

de modo que a matriz de D em relação à base ordenada $\mathscr{B}$ é

$$[D]_{\mathscr{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Vimos o que acontece às matrizes representantes quando as transformações são somadas, a saber, que as matrizes se somam. Gostaríamos agora de perguntar o que acontece quando compomos transformações. Mais especificamente, sejam V, W e Z espaços vetoriais sobre o corpo F, de dimensões n, m e p, respectivamente. Seja T uma transformação linear de V em W e U uma transformação linear de W em Z. Suponhamos que existam bases ordenadas

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\},\ \mathscr{B}' = \{\beta_1, \ldots, \beta_m\},\ \text{e}\ \mathscr{B}'' = \{\gamma_1, \ldots, \gamma_p\}$$

para os espaços V, W e Z, respectivamente. Seja A a matriz de T em relação ao par $\mathscr{B}$, $\mathscr{B}'$ e B a matriz de U em relação ao par $\mathscr{B}'$, $\mathscr{B}''$. É fácil ver então que a matriz C da transformação UT em relação ao par $\mathscr{B}$, $\mathscr{B}''$ é o produto de B por A, pois se α é um vetor arbitrário em V

$$[T\alpha]_{\mathscr{B}'} = A[\alpha]_{\mathscr{B}}$$
$$[U(T\alpha)]_{\mathscr{B}''} = B[T\alpha]_{\mathscr{B}'}$$

e então

$$[(UT)(\alpha)]_{\mathscr{B}''} = BA[\alpha]_{\mathscr{B}}$$

logo, pela definição e unicidade da matriz representante, temos, necessariamente, $C = BA$. Isto também pode ser visto efetuando os cálculos

$$\begin{aligned}(UT)(\alpha_j) &= U(T\alpha_j)\\ &= U\left(\sum_{k=1}^{m} A_{kj}\beta_k\right)\\ &= \sum_{k=1}^{m} A_{kj}(U\beta_k)\\ &= \sum_{k=1}^{m} A_{kj} \sum_{i=1}^{p} B_{ik}\gamma_i\\ &= \sum_{i=1}^{p}\left(\sum_{k=1}^{m} B_{ik}A_{kj}\right)\gamma_i\end{aligned}$$

de modo que temos

(3-6) $$C_{ij} = \sum_{k=1}^{m} B_{ik}A_{kj}.$$

A definição (3-6) de multiplicação de matrizes foi motivada por meio de operações sobre as linhas de uma matriz. Vê-se aqui que uma motivação bastante forte para a definição encontra-se na composição de transformações lineares. Resumindo formalmente:

Teorema 13. *Sejam* V, W *e* Z *espaços vetoriais de dimensão finita sobre o corpo* F; *seja* T *uma transformação linear de* V *em* W *e* U *uma transformação linear de* W *em* Z. *Se* $\mathscr{B}$, $\mathscr{B}'$ *e* $\mathscr{B}''$ *são bases ordenadas dos espaços* V, W *e* Z, *respectivamente, se* A *é matriz de* T *em relação ao par* $\mathscr{B}$, $\mathscr{B}'$ *e* B *é a matriz de* U *em relação ao par* $\mathscr{B}'$, $\mathscr{B}''$, *então a matriz da composta* UT *em relação ao par* $\mathscr{B}$, $\mathscr{B}'$ *é a matriz produto* C = BA.

Observamos que o Teorema 13 demonstra que a multiplicação de matrizes é associativa — uma demonstração que não requer cálculos e é independente da demonstração dada no Capítulo 1. Devemos destacar também que no Exemplo 12 demonstramos um caso especial do Teorema 13.

É importante notar que se T e U são operadores lineares sobre um espaço V e se estamos usando apenas uma base ordenada $\mathscr{B}$, então o Teorema 13 toma a forma simples $[UT]_{\mathscr{B}} = [U]_{\mathscr{B}}[T]_{\mathscr{B}}$ Assim, neste caso, a correspondência que $\mathscr{B}$ determina entre operadores lineares e matrizes é não somente um isomorfismo de espaço vetorial mas conserva também produtos. Uma conseqüência simples disto é que o operador linear T é inversível se, e somente se, $[T]_{\mathscr{B}}$ é uma matriz inversível. De fato, o operador idêntico I é representado pela matriz unidade em relação a qualquer base ordenada, portanto

$$UT = TU = I$$

é equivalente a

$$[U]_{\mathscr{B}}[T]_{\mathscr{B}} = [T]_{\mathscr{B}}[U]_{\mathscr{B}} = I.$$

Evidentemente, quando T é inversível

$$[T^{-1}]_{\mathscr{B}} = [T]^{-1}_{\mathscr{B}}.$$

Gostaríamos agora de perguntar o que acontece com as matrizes representantes quando mudamos a base ordenada. Para efeito de simplicidade, consideraremos esta questão apenas para operadores lineares sobre um espaço V, de modo que possamos usar uma única base ordenada. A questão específica é a seguinte: seja T um operador linear sobre o espaço de dimensão finita V e sejam a

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\} \quad \text{e} \quad \mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

duas bases ordenadas de V. Qual é a relação entre as matrizes $[T]_{\mathscr{B}}$ e $[T]_{\mathscr{B}'}$? Como observamos no Capítulo 2, existe uma única $n \times n$ matriz (inversível) P tal que

(3-7) $$[\alpha]_{\mathscr{B}} = P[\alpha]_{\mathscr{B}'}$$

para todo vetor α em V. É a matriz $P = [P_1, \ldots, P_n]$ onde $P_j = [\alpha'_j]_{\mathscr{B}}$. Por definição

(3-8) $$[T\alpha]_{\mathscr{B}} = [T]_{\mathscr{B}}[\alpha]_{\mathscr{B}}.$$

Aplicando (3-7) ao vetor $T\alpha$ temos

(3-9) $$[T\alpha]_{\mathscr{B}} = P[T\alpha]_{\mathscr{B}'}.$$

Combinando (3-7), (3-8) e (3-9), obtemos

$$[T]_{\mathscr{B}}P[\alpha]_{\mathscr{B}'} = P[T\alpha]_{\mathscr{B}'}$$

ou

$$P^{-1}[T]_{\mathscr{B}}P[\alpha]_{\mathscr{B}'} = [T\alpha]_{\mathscr{B}'}$$

e então é necessário que

(3-10) $$[T]_{\mathscr{B}'} = P^{-1}[T]_{\mathscr{B}}P.$$

Isto responde nossa pergunta.

Antes de enunciarmos formalmente este resultado, observemos um fato. Existe um único operador linear U que leva $\mathscr{B}$ sobre $\mathscr{B}'$ definido por

$$U\alpha_j = \alpha'_j, \quad j = 1, \ldots, n.$$

Este operador U é inversível uma vez que leva uma base de V sobre uma base de V. A matriz P (acima) é exatamente a matriz do operador U em relação à base ordenada $\mathscr{B}$. De fato, P é definida por

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i$$

e como $U\alpha_j = \alpha'_j$, esta equação pode ser escrita como

$$U\alpha_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Portanto $P = [U]_{\mathscr{B}}$, por definição.

Teorema 14. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F *e sejam*

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \quad e \quad \mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

bases ordenadas de V. *Suponhamos que* T *seja um operador linear sobre* V. *Se* $P = [P_1, \ldots, P_n]$ *é a* $n \times n$ *matriz com colunas* $P_j = [\alpha'_j]_{\mathcal{B}}$, *então*

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P.$$

Alternativamente, se U *é o operador inversível sobre* V *definido por* $U\alpha_j = \alpha'_j$, $j = 1, \ldots, n$, *então*

$$[T]_{\mathcal{B}'} = [U]_{\mathcal{B}}^{-1}[T]_{\mathcal{B}}[U]_{\mathcal{B}}.$$

Exemplo 16. Seja T o operador linear sobre R^2 definido por $T(x_1, x_2) = (x_1, 0)$. No Exemplo 14 mostramos que a matriz de T em relação à base ordenada canônica $\mathcal{B} = \{\varepsilon_1, \varepsilon_2\}$ é

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Suponhamos que $\mathcal{B}'$ seja a base ordenada de R^2 formada pelos vetores $\varepsilon'_1 = (1, 1)$, $\varepsilon'_2 = (2, 1)$. Então

$$\begin{aligned} \varepsilon'_1 &= \varepsilon_1 + \varepsilon_2 \\ \varepsilon'_2 &= 2\varepsilon_1 + \varepsilon_2 \end{aligned}$$

de modo que P é a matriz

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}.$$

Efetuando cálculos simples obtemos

$$P^{-1} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}.$$

Assim

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P$$

$$= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix}$$

Podemos verificar facilmente que isto está correto porque

$$\begin{aligned} T\varepsilon_1' &= (1, 0) = -\varepsilon_1' + \varepsilon_2' \\ T\varepsilon_2' &= (2, 0) = -2\varepsilon_1' + 2\varepsilon_2'. \end{aligned}$$

Exemplo 17. Seja V o espaço das funções polinomiais de R em R, de "grau" menor ou igual a 3. Como no Exemplo 15, seja D o operador derivação sobre V e seja

$$\mathscr{B} = \{f_1, f_2, f_3, f_4\}$$

a base ordenada de V definida por $f_i(x) = x^{i-1}$. Seja t um número real e definamos $g_i(x) = (x + t)^{i-1}$, isto é,

$$\begin{aligned} g_1 &= f_1 \\ g_2 &= tf_1 + f_2 \\ g_3 &= t^2 f_1 + 2tf_2 + f_3 \\ g_4 &= t^3 f_1 + 3t^2 f_2 + 3tf_3 + f_4. \end{aligned}$$

Como se pode ver facilmente, a matriz

$$P = \begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

é inversível com

$$P^{-1} = \begin{bmatrix} 1 & -t & t^2 & -t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

portanto decorre que $\mathcal{B}' = \{g_1, g_2, g_3, g_4\}$ é uma base ordenada de V. No Exemplo 15 ficamos sabendo que a matriz de D em relação à base ordenada $\mathcal{B}$ é

$$[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

A matriz de D em relação à base ordenada $\mathcal{B}'$ é portanto

$$P^{-1}[D]_{\mathcal{B}}P = \begin{bmatrix} 1 & -t & t^2 & t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -t & t^2 & t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 2 & 6t \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Assim D é representado pela mesma matriz em relação às bases ordenadas $\mathcal{B}$ e $\mathcal{B}'$. Evidentemente, isto pode ser visto um pouco mais diretamente pois

$$\begin{aligned} Dg_1 &= 0 \\ Dg_2 &= g_1 \\ Dg_3 &= 2g_2 \\ Dg_4 &= 3g_3. \end{aligned}$$

Este exemplo ilustra um fato interessante. Se se conhece a matriz de um operador linear em relação a alguma base ordenada $\mathcal{B}$ e quer-se determinar a matriz em relação a outra base ordenada $\mathcal{B}'$, freqüentemente o que mais convém é efetuar a mudança de coordenadas usando a matriz inversível P; contudo, pode ser muito mais fácil determinar a matriz representante recorrendo diretamente à sua definição.

Definição. *Sejam* A *e* B n × n *matrizes (quadradas) sobre o corpo* F. *Dizemos que* B *é* **semelhante a A sobre** F *se existe uma* n × n *matriz inversível* P *sobre* F *tal que* $B = P^{-1}AP$.

De acordo com o Teorema 14, temos o seguinte: Se V é um espaço vetorial n-dimensional sobre F e $\mathscr{B}$ e $\mathscr{B}'$ são duas bases ordenadas de V, então, para cada operador linear T sobre V, a matriz $B = [T]_{\mathscr{B}'}$, é semelhante à matriz $A = [T]_{\mathscr{B}}$. O argumento também vale no outro sentido. Suponhamos que A e B sejam $n \times n$ matrizes e que B seja semelhante a A. Seja V um espaço vetorial n-dimensional arbitrário sobre F e seja $\mathscr{B}$ uma base ordenada de V. Seja T o operador linear sobre V que é representado em relação à base $\mathscr{B}$ por A. Se $B = P^{-1}AP$, seja $\mathscr{B}'$ a base ordenada de V obtida de $\mathscr{B}$ por meio de P, isto é,

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Então, a matriz de T em relação à base ordenada $\mathscr{B}'$ será B.

Assim, a afirmação de que B é semelhante a A significa que em cada espaço n-dimensional sobre F as matrizes A e B representam a mesma transformação linear em relação a duas bases ordenadas (possivelmente) distintas.

Notemos que toda $n \times n$ matriz A é semelhante a si mesma, bastando tomar $P = I$; se B é semelhante a A, então A é semelhante a B, pois, $B = P^{-1}AP$ implica que $A = (P^{-1})^{-1}BP^{-1}$; se B é semelhante a A e C é semelhante a B, então C é semelhante a A, pois $B = P^{-1}AP$ e $C = Q^{-1}BQ$ implicam que $C = (PQ)^{-1}A(PQ)$. Assim, a semelhança é uma relação de equivalência sobre o conjunto das $n \times n$ matrizes sobre o corpo F. Notemos também que a única matriz semelhante à matriz unidade I é a própria I e que a única matriz semelhante à matriz nula é a própria matriz nula.

Exercícios

1. Seja T o operador linear sobre C^2 definido por $T(x_1, x_2) = (x_1, 0)$. Seja $\mathscr{B}$ a base ordenada canônica de C^2 e seja $\mathscr{B}' = \{\alpha_1, \alpha_2\}$ a base ordenada definida por $\alpha_1 = (1, i)$, $\alpha_2 = (-i, 2)$.
   (a) Qual é a matriz de T em relação ao par $\mathscr{B}$, $\mathscr{B}'$?
   (b) Qual é a matriz de T em relação ao par $\mathscr{B}'$, $\mathscr{B}$?
   (c) Qual é a matriz de T em relação à base ordenada $\mathscr{B}'$?
   (d) Qual é a matriz de T em relação à base ordenada $\{\alpha_2, \alpha_1\}$?

2. Seja T a transformação linear de R^3 em R^2 definida por

$$T(x_1, x_2, x_3) = (x_1 + x_2, 2x_3 - x_1).$$

(a) Se $\mathscr{B}$ é a base ordenada canônica de R^3 e $\mathscr{B}'$ é a base ordenada canônica de R^2, qual é a matriz de T em relação ao par $\mathscr{B}$, $\mathscr{B}'$?

(b) Se $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ e $\mathscr{B}' = \{\beta_1, \beta_2\}$, sendo

$$\alpha_1 = (1, 0, -1), \quad \alpha_2 = (1, 1, 1), \quad \alpha_3 = (1, 0, 0), \quad \beta_1 = (0, 1), \quad \beta_2 = (1, 0)$$

qual é a matriz de T em relação ao par $\mathscr{B}$, $\mathscr{B}'$?

3. Seja T um operador linear sobre F^n, seja A a matriz de T em relação à base ordenada canônica de F^n e seja W o subespaço de F^n gerado pelos vetores-colunas de A. Qual é a relação de W com T?

4. Seja V um espaço vetorial bidimensional sobre o corpo F e seja $\mathscr{B}$ uma base ordenada de V. Se T é um operador linear sobre V e

$$[T]_{\mathscr{B}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

demonstrar que $T^2 - (a + d)T + (ad - bc)I = 0$.

5. Seja T o operador linear sobre R^3, cuja matriz em relação à base ordenada canônica é

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 3 & 4 \end{bmatrix}.$$

Determinar uma base da imagem de T e uma base do núcleo de T.

6. Seja T o operador linear sobre R^2 definido por

$$T(x_1, x_2) = (-x_2, x_1).$$

(a) Qual é a matriz de T em relação à base ordenada canônica de R^2?

(b) Qual é a matriz de T em relação à base ordenada $\mathscr{B} = \{\alpha_1, \alpha_2\}$ sendo $\alpha_1 = (1, 2)$ e $\alpha_2 = (1, -1)$?

(c) Demonstrar que para todo número real c o operador $(T - cI)$ é inversível.

(d) Demonstrar que se $\mathscr{B}$ é uma base ordenada qualquer de R^2 e $[T]_{\mathscr{B}} = A$, então $A_{12}A_{21} \neq 0$.

7. Seja T o operador linear sobre R^3 definido por

$$T(x_1, x_2, x_3) = (3x_1 + x_3, -2x_1 + x_2, -x_1 + 2x_2 + 4x_3).$$

(a) Qual é a matriz de T em relação à base ordenada canônica de R^3?

(b) Qual é a matriz de T em relação à base ordenada

$$\{\alpha_1, \alpha_2, \alpha_3\}$$

sendo $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (-1, 2, 1)$, e $\alpha_3 = (2, 1, 1)$?

(c) Demonstrar que T é inversível e dar uma regra para T^{-1} como a que define T.

8. Seja θ um número real. Demonstrar que as duas matrizes seguintes são semelhantes sobre o corpo dos números complexos:

$$\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \quad \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$$

(*Sugestão*: Seja T o operador linear sobre C^2 que é representado pela primeira matriz em relação à base ordenada canônica. Determinar então vetores α_1 e α_2 tais que $T\alpha_1 = e^{i\theta}\alpha_1$, $T\alpha_2 = e^{-i\theta}\alpha_2$, e $\{\alpha_1, \alpha_2\}$ seja uma base.)

9. Seja V um espaço vetorial de dimensão finita sobre o corpo F e sejam S e T operadores lineares sobre V. Perguntamos: quando é que existem bases ordenadas $\mathscr{B}$ e $\mathscr{B}'$ de V tais que $[S]_{\mathscr{B}} = [T]_{\mathscr{B}'}$? Demonstrar que tais bases existem se, e somente se, existe um operador linear inversível U sobre V tal que $T = USU^{-1}$. (*Esboço de demonstração*: Se $[S]_{\mathscr{B}} = [T]_{\mathscr{B}''}$, seja U o operador que leva $\mathscr{B}$ sobre $\mathscr{B}'$; mostrar que $S = UTU^{-1}$. Reciprocamente se $T = USU^{-1}$ para algum U inversível, seja $\mathscr{B}$ uma base ordenada arbitrária de V e seja $\mathscr{B}'$ sua imagem por meio de U. Mostrar então que $[S]_{\mathscr{B}} = [T]_{\mathscr{B}'}$.)

10. Vimos que o operador linear T sobre R^2 definido por $T(x_1, x_2) = (x_1, 0)$ é representado em relação à base ordenada canônica pela matriz

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Este operador satisfaz $T^2 = T$. Demonstrar que se S é um operador linear sobre R^2 tal que $S^2 = S$, então $S = 0$ ou $S = I$ ou então existe uma base ordenada $\mathscr{B}$ de R^2 tal que $[S]_{\mathscr{B}} = A$ (acima).

11. Seja W o espaço das $n \times 1$ matrizes-colunas sobre um corpo F. Se A é uma $n \times n$ matriz sobre F, então A define um operador linear L_A sobre W por meio da multiplicação à esquerda: $L_A(X) = AX$. Demonstrar que todo operador linear sobre W é a multiplicação à esquerda por alguma $n \times n$ matriz, isto é, é L_A para algum A.
Suponhamos agora que V seja um espaço vetorial-dimensional sobre o corpo F e seja $\mathscr{B}$ uma base ordenada de V. Para cada α em V definamos $U\alpha = [\alpha]_{\mathscr{B}}$. Demonstrar que U é um isomorfismo de V em W. Se T é um operador linear sobre V; então UTU^{-1} é um operador linear sobre W. Isto significa que UTU^{-1} é a multiplicação à esquerda por alguma $n \times n$ matriz A. Qual é a matriz A?

12. Seja V um espaço vetorial n-dimensional sobre o corpo F e seja $\mathscr{B} = \{\alpha_1, \dots, \alpha_n\}$ uma base ordenada de V.

(a) De acordo com o Teorema 1, existe um único operador linear T sobre V tal que

$$T\alpha_j = \alpha_{j+1}, \quad j = 1, \ldots, n-1, \quad T\alpha_n = 0.$$

Qual é a matriz A de T em relação à base ordenada $\mathcal{B}$?

(b) Demonstrar que $T^n = 0$ mas $T^{n-1} \neq 0$.

(c) Seja S um operador linear arbitrário sobre V tal que $S^n = 0$ mas $S^{n-1} \neq 0$. Demonstrar que existe uma base ordenada $\mathcal{B}'$ de V tal que a matriz de S em relação à base ordenada $\mathcal{B}'$ é a matriz A da parte (a).

(d) Demonstrar que se M e N são $n \times n$ matrizes sobre F tais que $M^n = N^n = 0$ mas $M^{n-1} \neq 0 \neq N^{n-1}$, então M e N são semelhantes.

13. Sejam V e W espaços vetoriais de dimensão finita sobre o corpo F e seja T uma transformação linear de V em W. Se

$$\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\} \text{ e } \mathcal{B}' = \{\beta_1, \ldots, \beta_m\}$$

são bases ordenadas de V e W, respectivamente, definamos as transformações lineares $E^{p,q}$ como na demonstração do Teorema 5: $E^{p,q}(\alpha_i) = \delta_{iq}\beta_p$. Então as $E^{p,q}$, $1 \leq p \leq m$, $1 \leq q \leq n$, formam uma base de $L(V, W)$, e portanto

$$T = \sum_{p=1}^{m} \sum_{q=1}^{n} A_{pq} E^{p,q}$$

para certos escalares A_{pq} (as coordenadas de T em relação a esta base de $L(V, W)$). Mostrar que a matriz com elementos $A(p,q) = A_{pq}$ é exatamente a matriz de T em relação ao par $\mathcal{B}$, $\mathcal{B}'$.

3.5 Funcionais Lineares

Se V é um espaço vetorial sobre o corpo F, uma transformação linear f de V no corpo F dos escalares é também denominada **funcional linear** sobre V. Se partirmos do início, isto significa que f é uma função de V em F tal que

$$f(c\alpha + \beta) = cf(\alpha) + f(\beta)$$

para todos os vetores α e β em V e todos os escalares c em F. O conceito de funcional linear é importante nos estudos dos espaços de dimensão finita porque ajuda organizar e esclarecer a discussão de subespaços, equações lineares e coordenadas.

Exemplo 18. Seja F um corpo e sejam $a_1, \ldots, a_n$ escalares em F. Definamos a função f sobre F^n:

$$f(x_1, \ldots, x_n) = a_1x_1 + \ldots + a_nx_n.$$

Então f é um funcional linear sobre F^n, e é representado pela matriz $[a_1, \ldots, a_n]$ em relação à base ordenada canônica de F^n e à base $\{1\}$ de F:

$$a_j = f(\varepsilon_j), \quad j = 1, \ldots, n.$$

Todo funcional linear sobre F^n é dessa forma, para determinados escalares $a_1, \ldots, a_n$. Isto decorre imediatamente da definição de funcional linear pois colocamos $a_j = f(e_j)$ e usamos a linearidade

$$\begin{aligned} f(x_1, \ldots, x_n) &= f\left(\sum_j x_j \varepsilon_j\right) \\ &= \sum_j x_j f(\varepsilon_j) \\ &= \sum_j a_j x_j. \end{aligned}$$

Exemplo 19. Eis um exemplo importante de funcional linear. Seja n um inteiro positivo e F um corpo. Se A é uma $n \times n$ matriz com elementos em F, o **traço** de A é o escalar

$$\operatorname{tr} A = A_{11} + A_{22} + \ldots + A_{nn}.$$

A função traço é um funcional linear sobre o espaço das matrizes $F^{n \times n}$, pois

$$\begin{aligned} \operatorname{tr}(cA + B) &= \sum_{i=1}^{n} (cA_{ii} + B_{ii}) \\ &= c \sum_{i=1}^{n} A_{ii} + \sum_{i=1}^{n} B_{ii} \\ &= c \operatorname{tr} A + \operatorname{tr} B. \end{aligned}$$

Exemplo 20. Seja V o espaço vetorial das funções polinomiais do corpo F em si mesmo. Seja t um elemento de F. Se definirmos

$$L_t(p) = p(t)$$

então L_t será um funcional linear sobre V. Geralmente descreve-se isto dizendo que, para cada t, o cálculo do valor em t é um funcional linear sobre o espaço das funções polinomiais. Talvez devamos observar que o fato de as funções serem polinomiais, não desempenha nenhum papel neste exemplo. O cálculo do valor em t é um funcional linear no espaço de todas as funções de F em F.

Exemplo 21. Este talvez seja o funcional linear mais importante da matemática. Seja $[a, b]$ um intervalo fechado na reta real e seja $C([a, b])$ o espaço das funções contínuas definidas sobre $[a, b]$, com valores reais. Então

$$L(g) = \int_a^b g(t)\,dt$$

define um funcional linear L sobre $C([a, b])$.

Se V é um espaço vetorial, a coleção de todos os funcionais lineares sobre V constitui um espaço vetorial de uma maneira natural. Trata-se do espaço $L(V, F)$. Representamos este espaço por V^* e o denominamos **espaço dual** de V:

$$V^* = L(V, F).$$

Se V é de dimensão finita podemos obter uma descrição bastante explícita do espaço dual V^*. Pelo Teorema 5 sabemos alguma coisa sobre o espaço V^*, a saber, que

$$\dim V^* = \dim V.$$

Seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base de V. De acordo com o Teorema 1, existe (para cada i) um único funcional linear f_i sobre V tal que

(3-11) $$f_i(\alpha_j) = \delta_{ij}.$$

Desta maneira obtemos a partir de $\mathscr{B}$ um conjunto de n funcionais lineares distintos $f_1, \ldots, f_n$ sobre V. Estes funcionais também são linearmente independentes. De fato, suponhamos que

(3-12) $$f = \sum_{i=1}^{n} c_i f_i$$

Então

$$\begin{aligned} f(\alpha_j) &= \sum_{i=1}^{n} c_i f_i(\alpha_j) \\ &= \sum_{i=1}^{n} c_i \delta_{ij} \\ &= c_j. \end{aligned}$$

Em particular, se f é o funcional nulo, $f(\alpha_j) = 0$ para cada j, logo, os escalares c_j são todos nulos. Ora, $f_1, \ldots, f_n$ são n funcionais linearmente independentes e como sabemos que V^* tem dimensão n, devemos ter que $\mathscr{B}^* = \{f_1, \ldots, f_n\}$ é uma base de V^*. Esta base é dita a **base dual** de $\mathscr{B}$.

Teorema 15. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F *e seja* $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ *uma base de* V. *Então existe uma única base dual* $\mathscr{B}^* = \{f_1, \ldots, f_n\}$ *de* V^* *tal que* $f_i(\alpha_j) = \delta_{ij}$. *Para cada funcional linear* f *sobre* V *temos*

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i \tag{3-13}$$

e para cada vetor α *em* V *temos*

$$\alpha = \sum_{i=1}^{n} f_i(\alpha)\, \alpha_i. \tag{3-14}$$

Demonstração. Demonstraremos acima que existe uma única base que é "dual" a $\mathscr{B}$. Se f é um funcional linear sobre V então f é alguma combinação linear (3-12) dos f_i e, como observamos por (3-12), os escalares c_j são dados necessariamente por $c_j = f(\alpha_j)$. Analogamente, se

$$\alpha = \sum_{i=1}^{n} x_i \alpha_i$$

é um vetor em V, então

$$\begin{aligned} f_j(\alpha) &= \sum_{i=1}^{n} x_i f_j(\alpha_i) \\ &= \sum_{i=1}^{n} x_i \delta_{ij} \\ &= x_j \end{aligned}$$

de modo que a única expressão de α como combinação linear dos α_i é

$$\alpha = \sum_{i=1}^{n} f_i(\alpha)\, \alpha_i.$$

A equação (3-14) nos fornece agora uma maneira satisfatória de descrever o que é a base dual. Ela nos diz que se $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ é uma base ordenada de V e $\mathscr{B}^* = \{f_1, \ldots, f_n\}$ é a base dual, então f_i é exatamente a função que associa a cada vetor α em V a i-ésima coordenada de α em relação à base ordenada $\mathscr{B}$. Assim podemos também chamar as f_i de funções coordenadas para $\mathscr{B}$. A fórmula (3-13) quando combinada com (3-14) nos diz o seguinte: Se f está em V^* e $f(\alpha_i) = a_i$, então se

$$\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$$

temos

$$f(\alpha) = a_1x_1 + \ldots + a_nx_n. \tag{3-15}$$

Em outras palavras, se fixamos uma base ordenada $\mathscr{B}$ de V e descrevemos cada vetor em V por sua n-upla de coordenadas $(x_1, \ldots, x_n)$ em relação a $\mathscr{B}$, então, todo funcional linear sobre V é da forma (3-15). Esta é a generalização natural do Exemplo 18 onde foi visto o caso especial $V = F^n$ e $\mathscr{B} = \{\varepsilon_1, \ldots, \varepsilon_n\}$.

Exemplo 22. Seja V o espaço vetorial das funções polinomiais de R em R, de grau menor ou igual a 2. Sejam t_1, t_2 e t_3 três quaisquer números reais *distintos* e seja

$$L_i(p) = p(t_i).$$

Então L_1, L_2 e L_3 são funcionais lineares sobre V. Estes funcionais são linearmente independentes; de fato, suponhamos que

$$L = c_1L_1 + c_2L_2 + c_3L_3.$$

Se $L = 0$, isto é, se $L(p) = 0$ para todo p em V, então, aplicando L às "funções" polinomiais particulares 1, x, x^2, obtemos

$$\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ t_1c_1 + t_2c_2 + t_3c_3 &= 0 \\ t_1^2c_1 + t_2^2c_2 + t_3^2c_3 &= 0. \end{aligned}$$

Disto decorre que $c_1 = c_2 = c_3 = 0$, pois (como mostram alguns cálculos simples) a matriz

$$\begin{bmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \\ t_1^2 & t_2^2 & t_3^2 \end{bmatrix}$$

é inversível quando t_1, t_2 e t_3 são distintos. Portanto, os L_i são independentes e como V tem dimensão 3, estes funcionais formam uma base de V^*. Qual é a base de V cuja dual é esta? Tal base $\{p_1, p_2, p_3\}$ de V precisa satisfazer

$$L_i(p_j) = \delta_{ij}$$

ou

$$p_j(t_i) = \delta_{ij}.$$

Estas funções polinomiais são, como se vê facilmente,

$$p_1(x) = \frac{(x - t_2)(x - t_3)}{(t_1 - t_2)(t_1 - t_3)}$$

$$p_2(x) = \frac{(x - t_1)(x - t_3)}{(t_2 - t_1)(t_2 - t_3)}$$

$$p_3(x) = \frac{(x - t_1)(x - t_2)}{(t_3 - t_1)(t_3 - t_2)}$$

A base $\{p_1, p_2, p_3\}$ de V é interessante porque, de acordo com (3-14), temos, para cada p em V,

$$p = p(t_1)p_1 + p(t_2)p_2 + p(t_3)p_3.$$

Assim, se c_1, c_2 e c_3 são números reais arbitrários, existe exatamente uma função polinomial p sobre R de grau menor ou igual a 2 que satisfaz $p(t_j) = c_j$, $j = 1, 2, 3$. Esta função polinomial é $p = c_1p_1 + c_2p_2 + c_3p_3$.

Discutamos agora a relação entre funcionais lineares e subespaços. Se f é um funcional linear não nulo, o posto de f é 1, pois a imagem de f é um subespaço não nulo do corpo dos escalares e é necessariamente o corpo dos escalares. Se o espaço V subjacente for de dimensão finita, o teorema do posto mais a nulidade (Teorema 2) afirma que o núcleo N_f tem dimensão

$$\dim N_f = \dim V - 1.$$

Em um espaço vetorial de dimensão n, um subespaço de dimensão $n-1$ é denominado um **hiperplano.** Estes espaços são às vezes denominados subespaços de codimensão 1. Perguntamos: todo hiperplano é núcleo de um funcional linear? É fácil ver que a resposta é afirmativa. E não é muito mais difícil mostrar que todo subespaço d-dimensional de um espaço n-dimensional é a interseção dos núcleos de $(n-d)$ funcionais lineares (Teorema 16 abaixo).

Definição. *Se* V *é um espaço vetorial sobre o corpo* F *e* S *é um subconjunto de* V, *o* **anulador** *de* S *é o conjunto* S^0 *dos funcionais lineares* f *sobre tais* V *que* $f(\alpha)=0$ *para todo* α *em* S.

Deve ficar claro para o leitor que S^0 é um subespaço de V^*, seja S um subespaço de V ou não. Se S é o conjunto formado apenas pelo vetor nulo, então, $S^0=V^*$. Se $S=V$, então S^0 é o subespaço nulo de V^*. (Isto é fácil de ver para o caso em que V é de dimensão finita.)

Teorema 16. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F *e seja* W *um subespaço de* V. *Então*

$$\dim W + \dim W^0 = \dim V.$$

Demonstração. Seja k a dimensão de W e $\{\alpha_1, \ldots, \alpha_k\}$ uma base de W. Tomemos vetores $\alpha_{k+1}, \ldots, \alpha_n$ em V tais que $\{\alpha_1, \ldots, \alpha_n\}$ seja uma base de V. Seja $\{f_1, \ldots, f_n\}$ a base de V^* que é a dual desta base de V. Afirmamos que $\{f_{k+1}, \ldots, f_n\}$ é uma base do anulador W^0. Certamente f_i pertence a W^0 para $i \geq k+1$, porque

$$f_i(\alpha_j) = \delta_{ij}$$

e $\delta_{ij}=0$ se $i \geq k+1$ e $j \leq k$; disto decorre que, para $i \geq k+1$, $f_i(\alpha)=0$ sempre que α seja uma combinação linear de $\alpha_1, \ldots, \alpha_k$. Os funcionais $f_{k+1}, \ldots, f_n$ são independentes, portanto basta mostrar que eles geram W^0. Suponhamos que f esteja em V^*. Ora,

$$f = \sum_{i=1}^{n} f(\alpha_i) f_i$$

de modo que, se f está em W^0, temos $f(\alpha_i)=0$ para $i \leq k$ e

$$f = \sum_{i=k+1}^{n} f(\alpha_i) f_i.$$

Mostramos assim que se $\dim W = k$ e $\dim V = n$ então

$$\dim W^0 = n - k.$$

Corolário. *Se* W *é um subespaço* k-*dimensional de um espaço vetorial* V, n-*dimensional, então* W *é a interseção de* (n − k) *hiperplanos em* V.

Demonstração. Este é um corolário da demonstração do Teorema 16, mais do que do seu enunciado. Com a notação da demonstração, W é exatamente o conjunto dos vetores α tais que $f_i(\alpha) = 0$, $i = k + 1, \ldots, n$. Quando $k = n - 1$, W é o núcleo de f_n.

Corolário. *Se* W_1 *e* W_2 *são subespaços de um espaço vetorial de dimensão finita, então* $W_1 = W_2$ *se, e somente se,* $W_1^0 = W_2^0$.

Demonstração. Se $W_1 = W_2$ então certamente $W_1^0 = W_2^0$. Se $W_1 \neq W_2$ então um dos subespaços contém um vetor que não está no outro. Suponhamos que exista um vetor α que esteja em W_2 mas não em W_1. Pelos corolários anteriores (ou a demonstração do Teorema 16), existe um funcional linear f tal que $f(\beta) = 0$ para todo β em W, mas $f(\alpha) \neq 0$. Então f está em W_1 mas não em W_2^0 e $W_1^0 \neq W_2^0$.

Na próxima seção daremos uma demonstração diferente destes dois corolários. O primeiro corolário afirma que se fixarmos alguma base ordenada para o espaço, cada subespaço k-dimensional pode ser descrito, especificando-se $(n - k)$ condições lineares homogêneas sobre as coordenadas, em relação àquela base.

Consideremos rapidamente sistemas de equações lineares homogêneas do ponto de vista de funcionais lineares. Suponhamos ter um sistema de equações lineares

$$\begin{array}{c} A_{11}x_1 + \ldots + A_{1n}x_n = 0 \\ \vdots \\ A_{m1}x_1 + \ldots + A_{mn}x_n = 0 \end{array}$$

cujas soluções queiramos determinar. Se indicarmos por f_i, $i = 1, \ldots, m$, o funcional linear sobre F^n definido por

$$f_i(x_1, \ldots, x_n) = A_{i1}x_1 + \ldots + A_{in}x_n$$

então estamos procurando o subespaço de F^n constituído por todos α tais que

$$f_i(\alpha) = 0, \quad i = 1, \ldots, m.$$

Em outras palavras, estamos procurando o subespaço anulado por $f_1, \ldots, f_m$. A linha-redução da matriz dos coeficientes nos fornece um método sistemático para determinar esse subespaço. A n-upla $(A_{i1}, \ldots, A_{in})$ dá as coordenadas do funcional linear f_i em relação à base que é dual da base canônica de F^n. O espaço-linha da matriz dos coeficientes pode portanto ser considerado como o espaço de funcionais lineares gerado por $f_1, \ldots, f_m$. O espaço-solução é o subespaço anulado por esse espaço de funcionais.

Agora podemos considerar o sistema de equações do ponto de vista "dual". Isto é, suponhamos que nos sejam dados m vetores em F^n

$$\alpha_i = (A_{i1}, \ldots, A_{in})$$

e queiramos determinar o anulador do subespaço gerado por esses vetores. Como um funcional linear arbitrário sobre F^n tem a forma

$$f(x_1, \ldots, x_n) = c_1x_1 + \ldots + c_nx_n$$

a condição para que f esteja nesse anulador é que

$$\sum_{j=1}^{n} A_{ij}c_j = 0, \quad i = 1, \ldots, m$$

isto é, que $(c_1, \ldots, c_n)$ seja uma solução do sistema $AX = 0$. Sob este ponto de vista, a linha-redução nos dá um método sistemático para determinar o anulador do subespaço gerado por um dado conjunto finito de vetores em F^n.

Exemplo 23. Consideremos os três funcionais lineares sobre R^4:

$$\begin{aligned} f_1(x_1, x_2, x_3, x_4) &= x_1 + 2x_2 + 2x_3 + x_4 \\ f_2(x_1, x_2, x_3, x_4) &= 2x_2 + x_4 \\ f(x_1, x_2, x_3, x_4) &= -2x_1 - 4x_3 + 3x_4. \end{aligned}$$

O subespaço que eles anulam pode ser descrito explicitamente, determinando-se a matriz linha-reduzida à forma em escada que é linha-equivalente a

$$A = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ -2 & 0 & -4 & 3 \end{bmatrix}.$$

Um rápido cálculo ou um exame do Exemplo 21 do Capítulo 2, mostra que

$$R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Portanto os funcionais lineares

$$\begin{aligned} g_1(x_1, x_2, x_3, x_4) &= x_1 + 2x_3 \\ g_2(x_1, x_2, x_3, x_4) &= x_2 \\ g_3(x_1, x_2, x_3, x_4) &= x_4 \end{aligned}$$

geram o mesmo subespaço de $(R^4)^*$ e anulam o mesmo subespaço de R^4 que f_1, f_2, f_3. O subespaço anulado é constituído dos vetores com

$$\begin{aligned} x_1 &= -2x_3 \\ x_2 &= x_4 = 0 \end{aligned}$$

Exemplo 24. Seja W o subespaço de R^5 gerado pelos vetores

$$\alpha_1 = (2, -2, 3, 4, -1), \quad \alpha_3 = (0, 0, -1, -2, 3),$$
$$\alpha_2 = (-1, 1, 2, 5, 2), \quad \alpha_4 = (1, -1, 2, 3, 0).$$

Como se descreve W^0, o anulador de W? Formemos a 4×5 matriz A com vetores-linhas $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ e determinemos a matriz R linha-reduzida à forma em escada que é linha-equivalente a A;

$$A = \begin{bmatrix} 2 & -2 & 3 & 4 & -1 \\ -1 & 1 & 2 & 5 & 2 \\ 0 & 0 & -1 & -2 & 3 \\ 1 & -1 & 2 & 3 & 0 \end{bmatrix} \to R = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Se f é um funcional linear sobre R^5,

$$f(x_1, \ldots, x_5) = \sum_{j=1}^{5} c_j x_j$$

então f está em W^0 se, e somente se, $f(\alpha_i) = 0$, $i = 1, 2, 3, 4$, isto é, se, e somente se,

$$\sum_{j=1}^{5} A_{ij}c_j = 0, \quad 1 \leq i \leq 4.$$

Isto é equivalente a

$$\sum_{j=1}^{5} R_{ij}c_j = 0, \quad 1 \leq i \leq 3$$

ou

$$\begin{aligned} c_1 - c_2 - c_4 &= 0 \\ c_3 + 2c_4 &= 0 \\ c_5 &= 0. \end{aligned}$$

Obtemos todos esses funcionais lineares atribuindo valores arbitrários a c_2 e c_4, por exemplo, $c_2 = a$ e $c_4 = b$ e depois calculando os correspondentes $c_1 = a + b$, $c_3 = -2b$, $c_5 = 0$. Portanto, W^0 consiste de todos os funcionais lineares f da forma

$$f(x_1, x_2, x_3, x_4, x_5) = (a + b)x_1 + ax_2 - 2bx_3 + bx_4.$$

A dimensão de W^0 é 2 e pode-se encontrar uma base $\{f_1, f_2\}$ de W^0 tomando primeiro $a = 1$, $b = 0$ e depois $a = 0$, $b = 1$:

$$\begin{aligned} f_1(x_1, \ldots, x_5) &= x_1 + x_2 \\ f_2(x_1, \ldots, x_5) &= x_1 - 2x_3 + x_4. \end{aligned}$$

Um f arbitrário em W^0 é $f = af_1 + bf_2$.

Exercícios

1. Em R^3, sejam $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (0, 1, -2)$, $\alpha_3 = (-1, -1, 0)$.

   (a) Se f é um funcional linear sobre R^3 tal que

   $$f(\alpha_1) = 1, \quad f(\alpha_2) = -1, \quad f(\alpha_3) = 3,$$

   e se $\alpha = (a, b, c)$, determinar $f(\alpha)$.

   (b) Descrever explicitamente um funcional linear f sobre R^3 tal que

   $$f(\alpha_1) = f(\alpha_2) = 0 \text{ mas } f(\alpha_3) \neq 0.$$

(c) Seja f um funcional linear arbitrário tal que

$$f(\alpha_1) = f(\alpha_2) = 0 \text{ e } f(\alpha_3) \neq 0.$$

Se $\alpha = (2, 3, -1)$, mostrar que $f(\alpha) \neq 0$.

2. Seja $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3\}$ a base de C^3 definida por

$$\alpha_1 = (1, 0, -1), \quad \alpha_2 = (1, 1, 1), \quad \alpha_3 = (2, 2, 0).$$

Determinar a base dual de $\mathscr{B}$.

3. Se A e B são $n \times n$ matrizes sobre o corpo F, mostrar que traço (AB) = traço (BA). Mostrar depois que matrizes semelhantes têm o mesmo traço.

4. Seja V o espaço vetorial das funções polinomiais p de R em R que têm grau menor ou igual a 2, ou seja,

$$p(x) = c_0 + c_1 x + c_2 x^2.$$

Definamos três funcionais lineares sobre V por

$$f_1(p) = \int_0^1 p(x)\,dx \quad f_2(p) = \int_0^2 p(x)\,dx \quad f_3(p) = \int_0^{-1} p(x)\,dx.$$

Mostrar que $\{f_1, f_2, f_3\}$ é uma base de V^* exibindo a base de V da qual ela é dual.

5. Se A e B são $n \times n$ matrizes complexas, mostrar que é impossível ter-se $AB - BA = I$.

6. Sejam m e n inteiros positivos e F um corpo. Sejam $f_1, \ldots, f_m$ funcionais lineares sobre F^n. Para α em F^n definamos

$$T\alpha = (f_1(\alpha), \ldots, f_m(\alpha)).$$

Mostrar que T é uma transformação linear de F^n em F^m. Mostrar então que toda transformação linear de F^n em F^m é da forma acima, para certos $f_1, \ldots, f_m$.

7. Sejam $\alpha_1 = (1, 0, -1, 2)$ e $\alpha_2 = (2, 3, 1, 1)$ e seja W o subespaço de R^4 gerado por α_1 e α_2. Quais funcionais lineares f,

$$f(x_1, x_2, x_3, x_4) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4,$$

estão no anulador de W?

8. Seja W o subespaço de R^5 gerado pelos vetores

$$\alpha_1 = \varepsilon_1 + 2\varepsilon_2 + \varepsilon_3, \quad \alpha_2 = \varepsilon_2 + 3\varepsilon_3 + 3\varepsilon_4 + \varepsilon_5$$
$$\alpha_3 = \varepsilon_1 + 4\varepsilon_2 + 6\varepsilon_3 + 4\varepsilon_4 + \varepsilon_5.$$

Determinar uma base de W^0.

9. Seja V o espaço vetorial das 2×2 matrizes sobre o corpo dos números reais e seja

$$B = \begin{bmatrix} 2 & -2 \\ -1 & 1 \end{bmatrix}.$$

Seja W o subespaço de V que consiste de todas A tais que $AB = 0$. Seja f um funcional linear sobre V que esteja no anulador de W. Suponhamos que $f(I) = 0$ e $f(C) = 3$, sendo I a 2×2 matriz unidade e

$$C = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Determinar $f(B)$.

10. Seja F um subcorpo do corpo dos números complexos. Definamos n funcionais lineares sobre $F^n (n \geq 2)$ por

$$f_k(x_1, \ldots, x_n) = \sum_{j=1}^{n} (k - j) x_j, \quad 1 \leq k \leq n.$$

Qual é a dimensão do subespaço anulado por $f_1, \ldots, f_n$?

11. Sejam W_1 e W_2 subespaços de um espaço vetorial V de dimensão finita.

(a) Demonstrar que $(W_1 + W_2)^0 = W_1^0 \cap W_2^0$.

(b) Demonstrar que $(W_1 \cap W_2)^0 = W_1^0 + W_2^0$.

12. Seja V um espaço vetorial de dimensão finita sobre o corpo F e seja W um subespaço de V. Se f é um funcional linear sobre W, demonstrar que existe um funcional linear g sobre V tal que $g(\alpha) = f(\alpha)$ para todo α no subespaço W.

13. Seja F um subcorpo do corpo dos números complexos e seja V um espaço vetorial arbitrário sobre F. Suponhamos que f e g sejam funcionais lineares sobre V tais que a função h, definida por $h(\alpha) = f(\alpha) g(\alpha)$ também seja um funcional linear sobre V. Demonstrar que ou $f = 0$ ou $g = 0$.

14. Seja F um corpo de característica zero e seja V um espaço vetorial de dimensão finita sobre F. Se $\alpha_1, \ldots, \alpha_m$ são um número finito de vetores em V, todos diferentes do vetor nulo, demonstrar que existe um funcional linear f sobre V tal que

$$f(\alpha_i) \neq 0, \quad i = 1, \ldots, m.$$

15. De acordo com o Exercício 3, matrizes semelhantes possuem o mesmo traço. Portanto podemos definir o traço de um operador linear sobre um espaço de dimensão finita, como sendo o traço de qualquer matriz que represente o operador em relação a uma base ordenada. Essa é uma boa definição pois todas as matrizes representantes de um mesmo operador são semelhantes.

Seja agora V o espaço das 2×2 matrizes sobre o corpo F e seja P uma 2×2 matriz fixa. Seja T o operador linear sobre V definido por $T(A) = PA$. Demonstrar que traço $(T) = 2$ traço (P).

16. Mostrar que o funcional traço sobre $n \times n$ matrizes é o único no seguinte sentido: se W é o espaço das $n \times n$ matrizes sobre o corpo F e se f é um funcional linear sobre W tal que $f(AB) = f(BA)$ para todas as A e B em W, então f é um múltiplo escalar da função traço. Se, além disso, $f(I) = n$, então f é a função traço.

17. Seja W o espaço das $n \times n$ matrizes sobre o corpo F e seja W_0 o subespaço gerado pelas matrizes C da forma $C = AB - BA$.

Demonstrar que W_0 é exatamente o subespaço das matrizes que têm traço nulo. (*Sugestão*: Qual é a dimensão do espaço das matrizes de traço nulo? Usar as "matrizes unitárias", isto é, as matrizes com exatamente um elemento não-nulo, para construir um número suficiente de matrizes linearmente independentes da forma $AB - BA$.)

3.6 O Bidual

Uma das perguntas a respeito de bases duais que não foi respondida na última seção é se cada base de V^* é a dual de alguma base de V. Uma maneira de responder esta questão é considerar V^{**}, o espaço dual de V^*.

Se α é um vetor de V, ele induz um funcional linear L_α sobre V^*, definido por

$$L_\alpha(f) = f(\alpha), \; f \text{ em } V^*.$$

O fato de que L_α é linear é apenas uma reformulação da definição de operações lineares em V^*:

$$\begin{aligned} L_\alpha(cf + g) &= (cf + g)(\alpha) \\ &= (cf)(\alpha) + g(\alpha) \\ &= cf(\alpha) + g(\alpha) \\ &= cL_\alpha(f) + L_\alpha(g). \end{aligned}$$

Se V for de dimensão finita e $\alpha \neq 0$, então $L_\alpha \neq 0$; em outras palavras, existe um funcional linear f tal que $f(\alpha) \neq 0$. A demonstração é muito simples e foi dada na Seção 3.5: Escolhemos uma base ordenada $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V tal que $\alpha_1 = \alpha$ e para f, o funcional linear que associa a cada vetor em V sua primeira coordenada em relação a base ordenada $\mathscr{B}$.

Teorema 16. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F. *Para cada vetor* α *em* V *definamos*

$$L_\alpha(f) = f(\alpha), \ f \ em \ V^*.$$

A aplicação $\alpha \to L_\alpha$ *é então um isomorfismo de* V *em* V^{**}.

Demonstração. Mostramos que para todo α a função L_α é linear. Suponhamos que α e β estejam em V, c esteja em F e seja $\gamma = c\alpha + \beta$. Então, para todo f em V^*

$$\begin{aligned} L_\gamma(f) &= f(\gamma) \\ &= f(c\alpha + \beta) \\ &= cf(\alpha) + f(\beta) \\ &= cL_\alpha(f) + L_\beta(f), \end{aligned}$$

portanto,

$$L_\gamma = cL_\alpha + L_\beta.$$

Isto mostra que a aplicação $\alpha \to L_\alpha$ é uma transformação linear de V em V^{**}. Esta transformação é não-singular, pois de acordo com as observações acima $L_\alpha = 0$ se, e somente se, $\alpha = 0$. Portanto, $\alpha \to L_\alpha$ é uma transformação linear não-singular de V em V^{**}, e como

$$\dim V^{**} = \dim V^* = \dim V$$

o Teorema 9 nos diz que esta transformação é inversível, sendo portanto um isomorfismo de V em V^{**}.

Corolário. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F. *Se* L *é um funcional linear sobre o espaço* V^* *dual de* V, *então existe um único vetor* α *em* V *tal que*

$$L(f) = f(\alpha)$$

para todo f *em* V^*.

Corolário. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F. *Toda base de* V^* *é a dual de alguma base de* V.

Demonstração. Seja $\mathscr{B}^* = \{f_1, \ldots, f_n\}$ uma base de V^*. Pelo Teorema 15, existe uma base $\{L_1, \ldots, L_n\}$ de V^{**} tal que

$$L_i(f_j) = \delta_{ij}.$$

Usando o corolário acima, para cada i existe um vetor α_i em V tal que

$$L_i(f) = f(\alpha_i)$$

para toda f em V^*, isto é, tal que $L_i = L_{\alpha_i}$. Decorre imediatamente que $\{\alpha_1, \ldots, \alpha_n\}$ é uma base de V e que B^* é a dual desta base.

Em vista do Teorema 17, geralmente identificamos α com L_α e dizemos que V "é" o espaço dual de V^* ou que os espaços V e V^* são duais um do outro de uma maneira natural. Cada um é o espaço dual do outro. No último corolário vimos um exemplo de como este fato pode ser usado. Aqui está mais um exemplo:

Se E é um subconjunto de V^*, o anulador E^0 é (tecnicamente) um subconjunto de V^{**}. Se decidirmos identificar V e V^{**}, como no Teorema 17, então E^0 é um subespaço de V, a saber, o conjunto de todos os α em V tais que $f(\alpha) = 0$ para todo f em E. Num dos corolários do Teorema 16 observamos que cada subespaço W é determinado pelo seu anulador W^0. Determinado de que maneira? A resposta é que W é o subespaço anulado por todos os f em W^0, isto é, W é a interseção dos núcleos de todos os f em W^0. Como nossa atual notação para anuladores, a resposta pode ser formulada de uma maneira muito simples: $W = (W^0)^0$.

Teorema 18. *Se* S *é um subconjunto qualquer de um espaço vetorial* V *de dimensão finita, então* $(S^0)^0$ *é o subespaço gerado por* S.

Demonstração. Seja W o subespaço gerado por S. Obviamente $W^0 = S^0$. Portanto, devemos demonstrar que $W = W^{00}$. Já fizemos uma demonstração. Aqui está uma outra. Pelo Teorema 16

$$\dim W + \dim W^0 = \dim V$$
$$\dim W^0 + \dim W^{00} = \dim V^*$$

e sendo $\dim V = \dim V^*$, temos

$$\dim W = \dim W^{00}.$$

Como W é um subespaço de W^{00}, vemos que $W = W^{00}$.

Os resultados desta seção são válidos para espaços vetoriais arbitrários; no entanto, sua demonstração requer o uso do assim chamado Axioma da Escolha. Queremos evitar uma extensa

discussão deste axioma e portanto não trataremos de anuladores de espaços vetoriais em geral. Mas, existem dois resultados a respeito de funcionais lineares sobre espaços vetoriais arbitrários que, por serem tão fundamentais, deverão ser incluídos aqui.

Seja V um espaço vetorial. Queremos definir hiperplanos em V. A não ser que V seja de dimensão finita, não podemos dar essa definição através da dimensão do hiperplano. Mas, podemos traduzir a idéia de um espaço N, que preenche V a menos de uma dimensão, da seguinte maneira:

1. N é um subespaço próprio de V.

2. Se W é um subespaço de V contendo N, então ou $W = N$ ou $W = V$.

Tomadas em conjunto, as condições (1) e (2) dizem que N é um subespaço próprio e não existe nenhum subespaço próprio maior; em suma, N é um subespaço próprio maximal.

Definição. Se V *é um espaço vetorial, um* **hiperplano** *em* V *é um subespaço próprio maximal de* V.

Teorema 19. *Se* f *é um funcional linear não nulo sobre o espaço vetorial* V, *então o núcleo de* f *é* um hiperplano *em* V. *Reciprocamente, todo hiperplano de* V *é o núcleo de um (não único) funcional linear não nulo sobre* V.

Demonstração. Seja f um funcional linear não nulo sobre V e N_f o seu núcleo. Seja α um vetor em V que não esteja em N_f, isto é, um vetor tal que $f(\alpha) \neq 0$. Vamos mostrar que todo vetor de V pertence ao subespaço gerado por N_f e α. Este subespaço é formado por todos os vetores

$$\gamma + c\alpha \quad \gamma \text{ em } N_f,\ c \text{ em } F.$$

Seja β um vetor de V. Definamos

$$c = \frac{f(\beta)}{f(\alpha)}$$

o que faz sentido pois $f(\alpha) \neq 0$. Então todo vetor $\gamma = \beta - c\alpha$ está em N_f, pois

$$\begin{aligned} f(\gamma) &= f(\beta - c\alpha) \\ &= f(\beta) - cf(\alpha) \\ &= 0. \end{aligned}$$

Portanto, β pertence ao subespaço gerado por N_f e α.

Seja agora N um hiperplano em V. Fixemos um vetor α que não esteja em N. Como N é um subespaço próprio maximal, o subespaço gerado por N e α é o espaço todo V. Portanto, todo vetor β em V tem a forma

$$\beta = \gamma + c\alpha \quad \gamma \text{ em } N, \ c \text{ em } F.$$

O vetor γ e o escalar c são determinados de modo único por β. Se também tivéssemos

$$\beta = \gamma' + c'\alpha \qquad \gamma' \text{ em } N, \ c' \text{ em } F,$$

então

$$(c' - c)\alpha = \gamma - \gamma'.$$

Se $c' - c \neq 0$, α estaria em N; logo $c' = c$ e $\gamma' = \gamma$. Uma outra maneira de formular nossa conclusão é esta: Se β está em V, existe um único escalar c tal que $\beta - c\alpha$ está em N. Chamemos esse escalar de $g(\beta)$. É fácil ver que g é um funcional linear sobre V e que N é o núcleo de g.

Lema. *Se* f *e* g *são funcionais lineares sobre um espaço vetorial* V, *então* g *é um múltiplo escalar de* f *se, e somente se, o núcleo de* g *contiver o núcleo de* f, *isto é, se, e somente se,* $f(\alpha) = 0$ *implicar* $g(\alpha) = 0$.

Demonstração. Se $f = 0$ então também $g = 0$ e g é trivialmente um múltiplo escalar de f. Suponhamos $f \neq 0$, de modo que o núcleo N_f seja um hiperplano em V. Escolhamos um vetor α em V com $f(\alpha) \neq 0$ e seja

$$c = \frac{g(\alpha)}{f(\alpha)}.$$

O funcional linear $h = g - cf$ é nulo sobre N_f pois tanto f como g são nulos sobre N_f e $h(\alpha) = g(\alpha) - cf(\alpha)$. Assim h é nulo sobre o subespaço gerado por N_f e α — e este subespaço é V. Concluímos que $h = 0$, isto é, que $g = cf$.

Teorema 20. *Sejam* g, $f_1, \ldots, f_r$ *funcionais lineares sobre um espaço vetorial* V *e sejam* N, $N_1, \ldots, N_r$ *seus respectivos núcleos.*

Então g é uma combinação linear de $f_1, \ldots, f_r$ *se, e somente se,* N *contiver a interseção* $N_1 \cap \ldots \cap N_r$.

Demonstração. Se $g = c_1 f_1 + \ldots + c_r f_r$ e $f_i(\alpha) = 0$ para cada i, então obviamente $g(\alpha) = 0$. Portanto, N contém $N_1 \cap \ldots \cap N_r$.

Provaremos a recíproca (a parte "se" do teorema) por indução sobre o número r. O lema precedente demonstra o teorema para o caso $r = 1$. Suponhamos o resultado conhecido para $r = k - 1$ e sejam $f_1, \ldots, f_k$ funcionais lineares com núcleos $N_1, \ldots, N_k$ tais que $N_1 \cap \ldots \cap N_k$ esteja contido em N, o núcleo de g. Sejam $g', f'_1, \ldots, f'_{k-1}$ as restrições de $g, f_1, \ldots, f_{k-1}$ ao subespaço N_k. Então $g', f'_1, \ldots, f'_{k-1}$ são funcionais lineares sobre o espaço vetorial N_k. Além do mais, se α é um vetor em N_k e $f'_i(\alpha) = 0$, $i = 1, \ldots, k - 1$, então α pertence a $N_1 \cap \ldots \cap N_k$ e portanto $g'(\alpha) = 0$. Pela hipótese de indução (o caso $r = k - 1$), existem escalares c_i, tais que

$$g' = c_1 f'_1 + \ldots + c_{k-1} f'_{k-1}.$$

Seja

$$h = g - \sum_{i=1}^{k-1} c_i f_i. \tag{3-16}$$

Então h é um funcional linear sobre V e utilizando (3-16), $h(\alpha) = 0$ para todo α em N_k. Pelo lema anterior, h é um múltiplo escalar de f_k. Se $h = c_k f_k$, então

$$g = \sum_{i=1}^{k} c_i f_i.$$

Exercícios

1. Seja n um inteiro positivo e F um corpo. Seja W o conjunto de todos os vetores $(x_1, \ldots, x_n)$ em F^n tais que $x_1 + \ldots + x_n = 0$.

(a) Demonstrar que W^0 consiste dos funcionais lineares f da forma

$$f(x_1, \ldots, x_n) = c \sum_{j=1}^{n} x_j.$$

(b) Mostrar que o espaço W^* dual de W pode ser identificado de maneira "natural" com os funcionais lineares

$$f(x_1, \ldots, x_n) = c_1 x_1 + \ldots + c_n x_n$$

sobre F^n que satisfazem $c_1 + \ldots + c_n = 0$.

2. Usar o Teorema 20 para demonstrar o seguinte: Se W é um subespaço de dimensão finita de um espaço vetorial V e se $\{g_1, \ldots, g_r\}$ é uma base arbitrária de W^0, então

$$W = \bigcap_{i=1}^{r} N_{g_i}.$$

3. Sejam S um conjunto, F um corpo e $V(S;F)$ o espaço das funções de S em F:

$$(f + g)(x) = f(x) + g(x)$$
$$(cf)(x) = cf(x).$$

Seja W um subespaço n-dimensional arbitrário de $V(S;F)$. Mostrar que existem pontos $x_1, \ldots, x_n$ em S e funções $f_1, \ldots, f_n$ em W tais que $f_i(x_j) = \delta_{ij}$.

3.7 A Transposta de uma Transformação Linear

Suponhamos que V e W sejam dois espaços vetoriais sobre o corpo F e sejam uma transformação linear de V em W. Então T induz uma transformação linear de W^* em V^*, como segue. Supônhamos que g seja um funcional linear sobre W e seja

(3-17) $$f(\alpha) = g(T\alpha)$$

para cada α em V. Então (3-17) define uma função f de V em F, a saber, a composta de T, uma função de V em W, com g, uma função de W em F. Como T e g são ambas lineares, o Teorema 6 nos diz que f também é linear, isto é, f é um funcional linear sobre V. Assim T nos fornece uma função T^t que a cada funcional linear g sobre W faz corresponder um funcional linear $f = T^t g$ sobre V, definido por (3-17). Notemos também que T^t é na verdade uma transformação linear de W^* em V^*; de fato, se g_1 e g_2 estão em W^* e c é um escalar

$$\begin{aligned}[T^t(cg_1 + g_2)](\alpha) &= (cg_1 + g_2)(T\alpha)\\ &= cg_1(T\alpha) + g_2(T\alpha)\\ &= c(T^t g_1)(\alpha) + (T^t g_2)(\alpha)\end{aligned}$$

de modo que $T^t(cg_1 + g_2) = cT^t g_1 + T^t g_2$. Façamos um resumo.

Teorema 21. *Sejam* V *e* W *espaços vetoriais sobre o corpo* F. *Para cada transformação linear* T *de* V *em* W *existe uma única transformação linear* T^t *de* W^* *em* V^* *tal que*

$$(T^t g)(\alpha) = g(T\alpha)$$

para todos g *em* W^* *e* α *em* V.

Denominaremos T^t a **transposta** de T. Esta transformação T^t é freqüentemente denominada a adjunta de T; no entanto, não usaremos essa terminologia.

Teorema 22. *Sejam* V *e* W *espaços vetoriais sobre o corpo* F *e seja* T *uma transformação linear de* V *em* W. *O núcleo de* T^t *é o anulador da imagem de* T. *Se* V *e* W *são de dimensão finita, então*

(i) *posto* (T^t) = *posto* (T);

(ii) *a imagem de* T^t *é o anulador do núcleo de* T.

Demonstração. Se g está em W^*, então por definição

$$(T^t g)(\alpha) = g(T\alpha)$$

para todo α em V. A afirmação de que g está no núcleo de T^t significa que $g(T\alpha) = 0$ para todo α em V. Portanto, o núcleo de T^t é exatamente o anulador da imagem de T.

Suponhamos que V e W sejam de dimensão finita, digamos dim $V = n$ e dim $W = n$. (i) Seja r o posto de T, isto é, a dimensão da imagem de T. Pelo Teorema 16, o anulador da imagem de T tem dimensão $(m - r)$. Pela primeira afirmação deste Teorema, a nulidade de T^t deve ser $(m - r)$. Mas como T^t é uma transformação linear sobre um espaço vetorial m-dimensional, o posto de T é $m - (m - r) = r$, logo T e T^t têm o mesmo posto. (ii) Seja N o núcleo de T. Todo funcional que está na imagem de T^t está no anulador de N; de fato, suponhamos que $f = T^t g$ para algum g em W^*; então, se α está em N

$$f(\alpha) = (T^t g)(\alpha) = g(T\alpha) = g(0) = 0.$$

Ora, a imagem de T^t é um subespaço do espaço N^0, e

$$\dim N^0 = n - \dim N = \text{posto } (T) = \text{posto } (T^t)$$

de modo que a imagem de T^t deve ser exatamente N^0.

Teorema 23. *Sejam* V *e* W *espaços vetoriais de dimensão finita sobre o corpo* F. *Seja* $\mathscr{B}$ *uma base ordenada de* V *com base dual* $\mathscr{B}^*$ *e seja* $\mathscr{B}'$ *uma base ordenada de* W *com base dual* $\mathscr{B}'^*$. *Seja* T *uma transformação linear de* V *em* W; *seja* A *a matriz de* T *em relação a* $\mathscr{B}$, $\mathscr{B}'$ *e seja* B *a matriz de* T^t *em relação a* $\mathscr{B}'^*$, $\mathscr{B}^*$. *Então* $B_{ij} = A_{ji}$.

Demonstração. Sejam

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}, \qquad \mathscr{B}' = \{\beta_1, \ldots, \beta_m\},$$
$$\mathscr{B}^* = \{f_1, \ldots, f_n\}, \qquad \mathscr{B}'^* = \{g_1, \ldots, g_m\}.$$

Por definição,

$$T\alpha_j = \sum_{i=1}^{n} A_{ij}\beta_i, \qquad j = 1, \ldots, n$$

$$T^t g_j = \sum_{t=1}^{n} B_{ij} f_i, \qquad j = 1, \ldots, m.$$

Por outro lado,

$$\begin{aligned}
(T^t g_j)(\alpha_i) &= g_j(T\alpha_i) \\
&= g_j\left(\sum_{k=1}^{m} A_{ki}\beta_k\right) \\
&= \sum_{k=1}^{m} A_{ki} g_j(\beta_k) \\
&= \sum_{k=1}^{m} A_{ki}\delta_{jk} \\
&= A_{ji}.
\end{aligned}$$

Para qualquer funcional linear f sobre V

$$f = \sum_{i=1}^{m} f(\alpha_i) f_i.$$

Aplicando esta fórmula ao funcional $f = T^t g_j$ e usando o fato de que $(T^t g_j)(\alpha_i) = A_{ji}$ temos

$$T^t g_j = \sum_{i=1}^{n} A_{ji} f_i$$

do que decorre imediatamente que $B_{ij} = A_{ji}$.

Definição. *Se* A *é uma* m × n *matriz sobre o corpo* F, *a* **transposta** *de* A *é a* n × m *matriz* A^t *definida pr* $A^t_{ij} = A_{ji}$.

O Teorema 23 afirma, portanto, que se T é uma transformação linear de V em W, cuja matriz em relação a algum par de bases é A, então a transformação transposta T^t é representada em relação ao par de bases duais pela matriz transposta A^t.

Teorema 24. *Seja* A *uma* m × n *matriz arbitrária sobre o corpo* F. *Então o posto-linha de* A *é igual ao posto-coluna de* A.

Demonstração. Seja $\mathscr{B}$ a base ordenada canônica de F^n e $\mathscr{B}'$ a base ordenada canônica de F^m. Seja T a transformação linear de F^n em F^m tal que a matriz de T em relação ao par $\mathscr{B}$, $\mathscr{B}'$ seja A, isto é,

$$T(x_1, \ldots, x_n) = (y_1, \ldots, y_m)$$

onde

$$y_i = \sum_{j=1}^{n} A_{ij} x_j.$$

O posto-coluna de A é o posto da transformação T, pois a imagem de T consiste de todas as m-uplas que são combinações lineares dos vetores-colunas de A.

Em relação às bases duais $\mathscr{B}'^*$ e $\mathscr{B}^*$, a aplicação transposta T^t é representada pela matriz A^t. Como as colunas de A^t são as linhas de A, vemos, pelo mesmo raciocínio, que o posto-linha de A (o posto-coluna de A^t) é igual ao posto de T^t. Pelo Teorema 22, T e T^t têm o mesmo posto, logo o posto-linha de A é igual ao posto-coluna de A.

Vemos agora que se A é uma $m \times n$ matriz sobre F e T é a transformação linear de F^n em F^m definida acima, então

$$\text{posto } (T) = \text{posto-linha } (A) = \text{posto-coluna } (A)$$

e denominaremos este número simplesmente o **posto** de A.

Exemplo 25. Este exemplo será de natureza geral — mais uma discussão que um exemplo. Seja V um espaço vetorial n-dimensional sobre o corpo F e seja T um operador linear sobre V. Suponhamos que $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ seja uma base ordenada de V. A matriz de T em relação à base $\mathscr{B}$ é definida como sendo a $n \times n$ matriz A tal que

$$T\alpha_j = \sum_{j=1}^{m} A_{ij}\alpha_i;$$

em outras palavras, A_{ij} é a i-ésima coordenada do vetor $T\alpha_j$ em relação à base ordenada $\mathscr{B}$. Se $\{f_1, \ldots, f_n\}$ é a base dual de $\mathscr{B}$, isto pode ser enunciado simplesmente como

$$A_{ij} = f_i(T\alpha_j).$$

Vejamos o que acontece quando mudamos de base. Suponhamos que

$$\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$$

seja outra base ordenada de V, com base dual $\{f'_1, \ldots, f'_n\}$. Se B é a matriz de T em relação à base ordenada $\mathscr{B}'$, então

$$B_{ij} = f'_i(T\alpha'_j).$$

Seja U o operador linear inversível tal que $U\alpha_j = \alpha'_j$. Então, a transposta de U é dada por $U^t f'_i = f_i$. É fácil verificar que por ser U inversível, U^t também o é e $(U^t)^{-1} = (U^{-1})^t$. Assim $f'_i = (U^{-1})^t f_i$, $i = 1, \ldots, n$. Portanto,

$$\begin{aligned} B_{ij} &= [(U^{-1})^t f_i]\ (T\alpha'_j) \\ &= f_i(U^{-1}T\alpha'_j) \\ &= f_i(U^{-1}TU\alpha_j). \end{aligned}$$

O que significa isto? Bem, $f_i(U^{-1}TU\alpha_j)$ é o elemento i, j da matriz de $U^{-1}TU$ em relação à base ordenada $\mathscr{B}$. Nossos cálculos acima mostram que este escalar é também o elemento i, j da matriz de T em relação à base ordenada $\mathscr{B}'$. Em outras palavras,

$$\begin{aligned} [T]_{\mathscr{B}'} &= [U^{-1}TU]_{\mathscr{B}} \\ &= [U^{-1}]_{\mathscr{B}}[T]_{\mathscr{B}}[U]_{\mathscr{B}} \\ &= [U]_{\mathscr{B}}^{-1}[T]_{\mathscr{B}}[U]_{\mathscr{B}} \end{aligned}$$

e esta é exatamente a fórmula de mudança de base que deduzimos anteriormente.

Exercícios

1. Seja F um corpo e seja f o funcional linear sobre F^2 definido por $f(x_1, x_2) = ax_1 + bx_2$. Para cada um dos operadores lineares T seguintes, sendo $g = T^t f$, determinar $g(x_1, x_2)$.

(a) $T(x_1, x_2) = (x_1, 0)$;

(b) $T(x_1, x_2) = (-x_2, x_1)$;

(c) $T(x_1, x_2) = (x_1 - x_2, x_1 + x_2)$.

2. Seja V o espaço vetorial das funções polinomiais sobre o corpo dos números reais. Sejam a e b números reais fixos e seja f o funcional linear sobre V definido por

$$f(p) = \int_a^b p(x)\, dx.$$

Se D é o operador derivação sobre V, o que é $D^t f$?

3. Seja V o espaço das $n \times n$ matrizes sobre um corpo F e seja B uma $n \times n$ matriz fixa. Se T é o operador linear sobre V definido por $T(A) = AB - BA$ e se f é a função traço, o que é $T^t f$?

4. Seja V um espaço vetorial de dimensão finita sobre o corpo F e seja T um operador linear sobre V. Seja c um escalar e suponhamos que exista um vetor não-nulo α em V tal que $T\alpha = c\alpha$. Demonstrar que existe um funcional linear não-nulo f sobre V tal que $T^t f = cf$.

5. Seja A uma $m \times n$ matriz com elementos *reais*. Demonstrar que $A = 0$ se, e somente se, traço $(A^t A) = 0$.

6. Seja n um inteiro positivo e seja V o espaço das funções polinomiais sobre o corpo dos números reais, de grau menor ou igual a n, isto é, funções da forma

$$f(x) = c_0 + c_1 x + \ldots + c_n x^n.$$

Seja D o operador derivação sobre V. Determinar uma base do núcleo do operador transposto D^t.

7. Seja V um espaço vetorial de dimensão finita sobre o corpo F. Mostrar que $T \to T^t$ é um isomorfismo de $L(V, V)$ em $L(V^*; V^*)$.

8. Seja V o espaço vetorial das $n \times n$ matrizes sobre o corpo F.

(a) Se B é uma $n \times n$ matriz fixa, definamos uma função f_B sobre V por $f_B(A) =$ traço $(B^t A)$. Mostrar que f_B é um funcional linear sobre V.

(b) Mostrar que todo funcional linear sobre V é da forma acima, isto é, é f_B para algum B.

(c) Mostrar que $B \to f_B$ é um isomorfismo de V em V^*.

Capítulo 4
POLINÔMIOS

4.1 Álgebras

O objetivo deste capítulo é estabelecer algumas das propriedades básicas da álgebra dos polinômios sobre um corpo. A discussão será facilitada se introduzirmos primeiro o conceito de uma álgebra linear sobre um corpo.

Definição. *Seja* F *um corpo.* **Uma álgebra linear sobre o corpo** F *é um espaço vetorial* $\mathscr{A}$ *sobre* F *com uma operação adicional, dita* **multiplicação de vetores,** *que associa a cada par de vetores* α, β *em* $\mathscr{A}$ *um vetor* $\alpha\beta$ *em* $\mathscr{A}$ *dito o* **produto** *de* α *por* β *de maneira tal que*

(1) *a multiplicação é associativa,*

$$\alpha(\beta\gamma) = (\alpha\beta)\gamma$$

(2) *a multiplicação é distributiva em relação à adição,*

$$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma \quad e \quad (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$$

(3) *para cada escalar* c *em* F,

$$c(\alpha\beta) = (c\alpha)\beta = \alpha(c\beta).$$

Se existir um elemento 1 *em* $\mathscr{A}$ *tal que* $1\alpha = \alpha 1 = \alpha$ *para todo* α *em* $\mathscr{A}$, *denominaremos* $\mathscr{A}$ *uma* **álgebra linear com elemento unidade sobre** F *e denominaremos* 1 *o* **elemento unidade** *de* $\mathscr{A}$. *A álgebra* $\mathscr{A}$ *é dita* **comutativa** *se* $\alpha\beta = \beta\alpha$ *para todos* α *e* β *em* $\mathscr{A}$.

Exemplo 1. O conjunto das $n \times n$ matrizes sobre um corpo, com as operações usuais, é uma álgebra linear com elemento uni-

dade; em particular o próprio corpo é uma álgebra com elemento unidade. Esta álgebra não é comutativa se $n \geq 2$. O corpo é (evidentemente) comutativo.

Exemplo 2. O espaço dos operadores lineares sobre um espaço vetorial, com o composto como o produto, é uma álgebra linear com elemento unidade. Ela é comutativa se, e somente se, o espaço é unidimensional.

O leitor talvez tenha tido alguma experiência com o produto escalar e com o produto vetorial em R^3. Se é o caso, deve verificar que nenhum desses produtos é o do tipo descrito na definição de uma álgebra linear. O produto escalar, como o nome indica, associa cada par de vetores um escalar, e assim, certamente, não é o tipo de produto que ora discutimos. O produto vetorial associa de fato um vetor a cada par de vetores em R^3; no entanto, esta não é uma multiplicação associativa.

O restante desta seção será dedicado à construção de uma álgebra que é significativamente diferente das álgebras dos exemplos precedentes. Seja F um corpo e S o conjunto dos inteiros não-negativos. Pelo Exemplo 3 do Capítulo 2, o conjunto de todas as funções de S em F é um espaço vetorial sobre F. Indicaremos este espaço vetorial por F^∞. Os vetores em F^∞ são portanto seqüências infinitas $f = (f_0, f_1, f_2, \ldots)$ de escalares f_i em F. Se $g = (g_0, g_1, g_2, \ldots)$, com g_i em F e a, b são escalares em F, $af + bg$ é a seqüência infinita dada por

$$(4\text{-}1) \qquad af + bg = (af_0 + bg_0,\ af_1 + bg_1,\ af_2 + bg_2, \ldots).$$

Definamos um produto em F^∞ associado a cada par de vetores f, g em F^∞ o vetor fg que é dado por

$$(4\text{-}2) \qquad (fg)_n = \sum_{i=0}^{n} f_i g_{n-i}, \qquad n = 0, 1, 2, \ldots.$$

Assim

$$fg = (f_0 g_0,\ f_0 g_1 + f_1 g_0,\ f_0 g_2 + f_1 g_1 + f_2 g_0, \ldots)$$

e como

$$(gf)_n = \sum_{i=0}^{n} g_i f_{n-i} = \sum_{i=0}^{n} f_i g_{n-i} = (fg)_n$$

para $n = 0, 1, 2, \ldots$, segue que a multiplicação é comutativa, isto é, $fg = gf$. Se h também pertence a F^∞, então

$$\begin{aligned}[(fg)h]_n &= \sum_{i=0}^{n} (fg)_i h_{n-i} \\ &= \sum_{i=0}^{n} \left(\sum_{j=0}^{i} f_j g_{i-j} \right) h_{n-i} \\ &= \sum_{i=0}^{n} \sum_{j=0}^{i} f_j g_{i-j} h_{n-i} \\ &= \sum_{j=0}^{n} f_j \sum_{i=0}^{n-j} g_i h_{n-i-j} \\ &= \sum_{j=0}^{n} f_j (gh)_{n-j} = [f(gh)]_n\end{aligned}$$

para $n = 0, 1, 2, \ldots$, segue que a multiplicação é comutativa, isto é,

$$(fg)h = f(gh). \tag{4-3}$$

Deixamos a cargo do leitor verificar que a multiplicação definida por (4-2) satisfaz (2) e (3) da definição de uma álgebra linear e que o vetor $1 = (1, 0, 0, \ldots)$ funciona como um elemento unidade de F^∞. Então, F^∞, com as operações definidas acima, é uma álgebra linear comutativa com elemento unidade sobre o corpo F.

O vetor $(0, 1, 0, \ldots, 0, \ldots)$ desempenha um papel notável no que segue e indicá-lo-emos consistentemente por x. Em todo este capítulo, x nunca será usado para indicar um elemento do corpo F. O produto de x por si mesmo n vezes será indicado por x^n e colocaremos $x^0 = 1$. Então

$$x^2 = (0, 0, 1, 0, \ldots), \qquad x^3 = (0, 0, 0, 1, 0, \ldots)$$

e, de maneira geral, para cada inteiro $k \geq 0$, $(x^k)_k = 1$ e $(x^k)_n = 0$ para todos inteiros não-negativos $n \neq k$. Concluindo esta seção, observemos que o conjunto formado por $1, x, x^2, \ldots$ é independente e infinito. Assim a álgebra F^∞ não é de dimensão finita.

A álgebra F^∞ é denominada, algumas vezes, álgebra das séries formais de potências sobre F. O elemento $f = (f_0, f_1, f_2, \ldots)$ é freqüentemente representado por

$$f = \sum_{n=0}^{\infty} f_n x^n. \qquad (4\text{-}4)$$

Esta notação é muito conveniente para efetuar operações algébricas. Quando ela for usada, deve se ter em mente que ela é puramente formal. Não existem "somas infinitas" em álgebra e a notação (4-4) de séries de potências não deve sugerir nada quanto a convergência, se é que o leitor sabe o que seja isto. Usando seqüências, foi possível definir precisamente uma álgebra na qual as operações se comportam como adição e multiplicação de séries formais de potências, sem correr o risco de confusão a respeito de objetos como somas infinitas.

4.2 A Álgebra dos Polinômios

Estamos agora em condições de definir um polinômio sobre o corpo F.

Definição. *Seja* F[x] *o subespaço de* F^{∞} *gerado pelos vetores* 1, x, x^2, ... *Um elemento de* F[x] *é dito um* **polinômio sobre** F.

Como $F[x]$ consiste de todas as combinações lineares (finitas) de x e suas potências, um vetor não-nulo f em F^{∞} é um polinômio se, e somente se, existe um inteiro $n \geq 0$ tal que $f_n \neq 0$ e tal que $f_k = 0$ para todos os inteiros $k > n$; este inteiro (quando existe) é evidentemente único e é denominado o **grau** de f. Indicamos o grau de um polinômio por $\mathrm{gr}(f)$ e não atribuímos nenhum grau ao polinômio nulo. Se f é um polinômio não-nulo de grau n temos que

$$f = f_0 x^0 + f_1 x + f_2 x^2 + \ldots + f_n x^n,\ f_n \neq 0. \qquad (4\text{-}5)$$

Os escalares $f_0, f_1, \ldots, f_n$ são às vezes ditos os **coeficientes** de f e podemos dizer que f é um polinômio com coeficientes em F. Denominaremos polinômios da forma cx^0 **polinômios constantes** e freqüentemente indicaremos xc^0 por c. Um polinômio não-nulo f de grau n tal que $f_n = 1$ é dito um polinômio **unitário.**

O leitor deve notar que polinômios não são a mesma coisa que funções polinomiais sobre F, já discutidas em várias ocasiões. Se F possuir infinitos elementos, existe um isomorfismo natural entre $F[x]$ e a álgebra das funções polinomiais sobre F. Discuti-

remos isto na próxima seção. Verifiquemos que $F[x]$ é uma álgebra.

Teorema 1. *Sejam* f *e* g *polinômios não-nulos sobre* F. *Então*

(i) fg *é um polinômio não-nulo*;

(ii) $gr(\mathrm{fg}) = gr(\mathrm{f}) + gr(\mathrm{g})$;

(iii) fg *é um polinômio unitário se* f *e* g *são ambos polinômios unitários*,

(iv) fg *é um polinômio constante se, e somente se,* f *e* g *são polinômios constantes*;

(v) *se* $\mathrm{f} + \mathrm{g} \neq 0$,

$$gr(\mathrm{f} + \mathrm{g}) \leq \textit{máx}\,(gr(\mathrm{f}),\ gr(\mathrm{g})).$$

Demonstração. Suponhamos que f tenha grau m e que g tenha grau n. Se k é um inteiro não-negativo.

$$fg)_{m+n+k} = \sum_{i=0}^{m+n+k} f_i g_{m+n+k-i}.$$

Para que $f_i g_{m+n+k-i} \neq 0$, é necessário que $i \leq m$ e $m + n + k - i \leq n$. Logo, é necessário que $m + k \leq i \leq m$, o que implica $k = 0$ e $i = m$. Assim,

$$(fg)_{m+n} = f_m g_n \tag{4-6}$$

e

$$(fg)_{m+n+k} = 0, \qquad k > 0. \tag{4-7}$$

As afirmações (i), (ii), (iii) decorrem imediatamente de (4-6) e (4-7), enquanto (iv) é uma conseqüência de (i) e (ii). Deixamos a verificação de (v) a cargo do leitor.

Corolário 1. *O conjunto dos polinômios sobre um dado corpo* F *é uma álgebra linear comutativa com elemento unidade sobre* F, *em relação às operações dadas por* (4-1) *e* (4-2).

Demonstração. Como as operações (4-1) e (4-2) são aquelas definidas na álgebra F^∞ e como $F[x]$ é um subespaço de F^∞,

basta demonstrar que o produto de dois polinômios é novamente um polinômio. Isto é trivial quando um dos fatores é nulo e, em caso contrário, decorre de (i).

Corolário 2. *Suponhamos que* f, g *e* h *sejam polinômios sobre o corpo* F *tais que* $f \neq 0$ *e* $fg = fh$. *Então* $g = h$.

Demonstração. Como $fg = fh$, $f(g - h) = 0$, e como $f \neq 0$ decorre imediatamente de (i) que $g - h = 0$.

Certos fatos adicionais decorrem bastante facilmente da demonstração do Teorema 1 e mencionaremos alguns deles.

Suponhamos que

$$f = \sum_{i=0}^{m} f_i x^i \quad \text{e} \quad g = \sum_{j=0}^{n} g_j x^j.$$

Então, de (4-7) obtemos

$$fg = \sum_{s=0}^{m+n} \left(\sum_{r=0}^{s} f_r g_{s-r} \right) x^s. \tag{4-8}$$

O leitor deve verificar que, no caso particular de $f = cx^m$, $g = dx^n$ com c, d em F, (4-8) reduz-se a

$$(cx^m)(dx^n) = cdx^{m+n}. \tag{4-9}$$

Ora, de (4-9) e das leis distributivas em $F[x]$, segue que o produto em (4-8) também é dado por

$$\sum_{i,j} f_i g_j x^i x^j \tag{4-10}$$

onde a soma é estendida a todos os pares i, j de inteiros tais que $0 \leq i \leq m$, e $0 \leq j \leq n$.

Definição. *Seja* $\mathscr{A}$ *uma álgebra linear com elemento unidade sobre o corpo* F. *Indicaremos o elemento unidade de* $\mathscr{A}$ *por* 1 *e faremos a convenção* $\alpha^0 = 1$ *para todo* α *em* $\mathscr{A}$. *Então, a cada polinômio* $f = \sum_{i=0}^{n} f_i x^i$ *sobre* F *e* α *em* $\mathscr{A}$, *associamos um elemento* $f(\alpha)$ *em* $\mathscr{A}$ *pela regra*

$$f(\alpha) = \sum_{i=0}^{n} f_i \alpha^i.$$

Exemplo 3. Seja C o corpo dos números complexos e seja $f = x^2 + 2$.

(a) Se $\mathscr{A} = C$ e z pertence a C, $f(z) = z^2 + 2$, em particular $f(2) = 6$ e

$$f\left(\frac{1+i}{1-i}\right) = 1.$$

(b) Se $\mathscr{A}$ é a álgebra das 2×2 matrizes sobre C e se

$$B = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$$

então

$$f(B) = 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}^2 = \begin{bmatrix} 3 & 0 \\ -3 & 6 \end{bmatrix}.$$

(c) Se $\mathscr{A}$ é a álgebra dos operadores lineares sobre C^3 e T é o elemento de $\mathscr{A}$ dado por

$$T(c_1, c_2, c_3) = (i\sqrt{2}\, c_1, c_2, i\sqrt{2}\, c_3)$$

então $f(T)$ é o operador linear sobre C^3 definido por

$$f(T)(c_1, c_2, c_3) = (0, 3c_2, 0).$$

(d) Se $\mathscr{A}$ é a álgebra dos polinômios sobre C e $g = x^4 + 3i$, então $f(g)$ é o polinômio em $\mathscr{A}$ dado por

$$f(g) = -7 + 6ix^4 + x^8.$$

O leitor observador poderá notar em relação a este último exemplo que, se f é um polinômio sobre um corpo arbitrário e x é o polinômio $\{0, 1, 0, \ldots\}$, então $f = f(x)$, mas aconselhamo-lo a esquecer este fato.

Teorema 2. *Seja* F *um corpo e* $\mathcal{A}$ *uma álgebra linear com elemento unidade sobre* F. *Suponhamos que* f *e* g *sejam polinômios sobre* F, *que* α *seja um elemento de* $\mathcal{A}$ *e que* c *pertença a* F. *Então*

(i) $(cf + g)(\alpha) = cf(\alpha) + g(\alpha)$;

(ii) $(fg)(\alpha) = f(\alpha)\,g(\alpha)$.

Demonstração. Como (i) é bem fácil de demonstrar, demonstraremos somente (ii).

Suponhamos que

$$f = \sum_{i=0}^{m} f_i x^i \quad \text{e} \quad g = \sum_{j=0}^{n} g_j x^j.$$

Por (4-10)

$$fg = \sum_{i,j} f_i g_j x^{i+j}$$

e, portanto, por (i)

$$\begin{aligned}(fg)(\alpha) &= \sum_{i,j} f_i g_j \alpha^{i+j}\\ &= \left(\sum_{i=0}^{m} f_i \alpha^i\right)\left(\sum_{j=0}^{n} g_j \alpha^j\right)\\ &= f(\alpha)\,g(\alpha).\end{aligned}$$

Exercícios

1. Seja F um subcorpo do corpo dos números complexos e seja A a seguinte 2×2 matriz sobre F:

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}.$$

Para cada um dos seguintes polinômios f sobre F, calcular $f(A)$:

(a) $f = x^2 - x + 2$;

(b) $f = x^3 - 1$;

(c) $f = x^2 - 5x + 7$.

2. Seja T o operador linear sobre R^3 definido por

$$T(x_1, x_2, x_3) = (x_1, x_3, -2x_2 - x_2).$$

Seja f o polinômio sobre R definido por $f = -x^3 + 2$. Determinar $f(T)$.

3. Seja A uma $n \times n$ matriz diagonal sobre o corpo F, isto é, uma matriz que satisfaz $A_{ij} = 0$ para $i \neq j$. Seja f o polinômio sobre F definido por

$$f = (x - A_{11}) \ldots (x - A_{nn}).$$

Qual é a matriz $f(A)$?

4. Se f e g são polinômios independentes sobre um corpo F e h é um polinômio não-nulo sobre F, mostrar que fh e gh são independentes.

5. Se F é um corpo, mostrar que o produto de dois elementos não-nulos de F^∞ é não-nulo.

6. Seja S um conjunto de polinômios não-nulos sobre um corpo F. Se dois quaisquer elementos de F nunca têm o mesmo grau, mostrar que S é um conjunto independente em $F[x]$.

7. Se a e b são elementos de um corpo F e $a \neq 0$, mostrar que os polinômios 1, $ax + b$, $(ax + b)^2$, $(ax + b)^3$,... formam uma base de $F[x]$.

8. Se F é um corpo e h é um polinômio sobre F de grau ≥ 1, mostrar que a aplicação $f \to f(h)$ é uma transformação linear injetora de $F[x]$ em $F[x]$. Mostrar que esta transformação é um isomorfismo de $F[x]$ em $F[x]$ se, e somente se, $\text{gr}(h) = 1$.

9. Seja F um subcorpo do corpo dos números complexos e sejam T e D transformações sobre $F[x]$ definidas por

$$T\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=0}^{n} \frac{c_i}{1+i} x^{i+1},$$

e

$$D\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=1}^{n} i c_i x^{i-1}.$$

(a) Mostrar que T é um operador linear não-singular sobre $F[x]$. Mostrar também que T não é inversível.

(b) Mostrar que D é um operador linear sobre $F[x]$ e determinar seu núcleo.

(c) Mostrar que $DT = I$ e $TD \neq I$.

(d) Mostrar que $T[(Tf)g] = (Tf)(Tg) - T[f(Tg)]$ para todos f e g em $F[x]$.

(e) Enunciar e demonstrar uma regra para D semelhante à regra dada para T em (*d*).

(f) Suponhamos que V seja um subespaço não-nulo de $F[x]$ tal que Tf pertença a V para todo f em V. Mostrar que V não é de dimensão finita.

(g) Suponhamos que V seja um subespaço de $F[x]$ que tenha dimensão finita. Demonstrar que existe um inteiro $m \geq 0$ tal que $D^m f = 0$ para todo f em V.

4.3 Interpolação de Lagrange

Em toda esta seção suporemos que F seja um corpo fixo e que $t_0, t_1, \ldots, t_n$ sejam $n+1$ elementos *distintos* de F. Seja V o subespaço de $F[x]$ que consiste dos polinômios de grau menor ou igual a n (mais o polinômio nulo) e seja L_i a função de V em F definida para f em V por

$$L_i(f) = f(t_i), \qquad 0 \leq i \leq n.$$

Pela parte (i) do Teorema 2, cada L_i é um funcional linear sobre V e uma das coisas que pretendemos mostrar é que o conjunto formado por $L_0, L_1, \ldots, L_n$ é uma base do espaço V^* dual de V.

Evidentemente, para que isto ocorra, é necessário e suficiente (cf. Teorema 15 do Capítulo 3) que $\{L_0, L_1, \ldots, L_n\}$ seja a dual de uma base $\{P_0, P_1, \ldots, P_n\}$ de V. Existe no máximo uma tal base e, se existe, é caracterizada por

$$L_j(P_i) = P_i(t_j) = \delta_{ij}. \tag{4-11}$$

Os polinômios

$$P_i = \frac{(x - t_0) \ldots (x - t_{i-1})(x - t_{i+1}) \ldots (x - t_n)}{(t_i - t_0) \ldots (t_i - t_{i-1})(t_i - t_{i+1}) \ldots (t_i - t_n)} \tag{4-12}$$

$$= \prod_{j \neq i} \left(\frac{x - t_j}{t_i - t_j} \right)$$

são de grau n, logo pertencem a V e, pelo Teorema 2, satisfazem (4-10).

Se $f = \sum_i c_i P_i$, então, para cada j,

$$f(t_j) = \sum_i c_i P_i(t_j) = c_j. \tag{4-13}$$

Como o polinômio nulo tem a propriedade de que $0(t) = 0$ para todo t em F, decorre de (4-13) que os polinômios $P_0, P_1, \ldots, P_n$ são linearmente independentes. Os polinômios $1, x, \ldots, x^n$ formam uma base de V, logo a dimensão de V é $n+1$. Portanto o conjunto independente $\{P_0, P_1, \ldots, P_n\}$ deve também ser uma base de V. Assim, para todos f em V

$$f = \sum_{i=0}^{n} f(t_i) P_i. \tag{4-14}$$

A expressão (4-14) é denominada a **fórmula de interpolação de Lagrange.** Tomando $f = x^j$ em (4-14) obtemos

$$x^j = \sum_{i=0}^{n} (t_i)^j P_i.$$

Decorre então do Teorema 7 do Capítulo 2 que a matriz

$$\text{(4-15)} \qquad \begin{bmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^n \\ 1 & t_1 & t_1^2 & \cdots & t_1^n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & t_n & t_n^2 & \cdots & t_n^n \end{bmatrix}$$

é inversível. A matriz em (4-15) é dita uma **matriz de Vandermonde;** constitui um exercício interessante mostrar diretamente que uma tal matriz é inversível, quando $t_0, t_1, \ldots, t_n$ são $n+1$ elementos distintos de F.

Se f é um polinômio arbitrário sobre F, indicaremos por $f^{\sim}$ em nossa presente discussão a função polinominal de F em F que leva cada t em F em $f(t)$. Por definição (cf. Exemplo 4 do Capítulo 2) toda função polinomial surge desta maneira; contudo, pode acontecer que $f^{\sim} = g^{\sim}$ para dois polinômios f e g tais que $f \neq g$. Felizmente, como veremos, esta situação desagradável ocorre apenas quando F é um corpo com um número finito de elementos distintos. Para descrever de maneira precisa a relação entre polinômios e funções polinomiais, precisamos definir o produto de duas funções polinomiais. Se f e g são polinômios sobre F, o produto de $f^{\sim}$ por $g^{\sim}$ é a função $f^{\sim}g^{\sim}$ de F em F dada por

$$\text{(4-16)} \qquad (f^{\sim}g^{\sim})(t) = f^{\sim}(t)g^{\sim}(t), \qquad t \text{ em } F.$$

Pela parte (ii) do Teorema 2, $(fg)(t) = f(t)g(t)$, logo

$$(fg)^{\sim}(t) = f^{\sim}(t)\,g^{\sim}(t)$$

para cada t em F. Assim, $f^{\sim}g^{\sim} = (fg)^{\sim}$ e é uma função polinomial. Neste ponto, um fato de verificação imediata, que deixamos a cargo do leitor, é que o espaço vetorial das funções polinomiais sobre F torna-se uma álgebra linear com elemento unidade sobre F se a multiplicação é definida por (4-16).

Definição. *Seja* F *um corpo e sejam* $\mathscr{A}$ *e* $\mathscr{A}^\sim$ *álgebras lineares sobre* F. *As álgebras* $\mathscr{A}$ *e* $\mathscr{A}^\sim$ *são ditas isomorfas se existe uma aplicação bijetora* $\alpha \to \alpha^\sim$ *de* $\mathscr{A}$ *em* $\mathscr{A}^\sim$ *tal que*

(a) $$(c\alpha + d\beta)^\sim = c\alpha^\sim + d\beta^\sim$$

(b) $$(\alpha\beta)^\sim = \alpha^\sim\beta^\sim$$

para todos α *e* β *em* $\mathscr{A}$ *e todos escalares* c *e* d *em* F. *A aplicação* $\alpha \to \alpha^\sim$ *é dita um isomorfismo de* $\mathscr{A}$ *em* $\mathscr{A}^\sim$. *Um isomorfismo de* $\mathscr{A}$ *em* $\mathscr{A}^\sim$ *é assim um isomorfismo de espaço vetorial de* $\mathscr{A}$ *em* $\mathscr{A}^\sim$ *que tem a propriedade adicional* (b) *de "conservar" produtos.*

Exemplo 4. Seja V um espaço vetorial n-dimensional sobre o corpo F. Pelo Teorema 13 do Capítulo 3 e observações subseqüentes, cada base ordenada $\mathscr{B}$ de V determina um isomorfismo $T \to [T]_{\mathscr{B}}$ da álgebra dos operadores lineares sobre V na álgebra das $n \times n$ matrizes sobre F. Suponhamos agora que U seja um operador linear fixo sobre V e que nos seja dado um polinômio

$$f = \sum_{i=0}^{n} c_i x^i$$

com coeficientes c_i em F. Então

$$f(U) = \sum_{i=0}^{n} c_i U^i$$

e como $T \to [T]_{\mathscr{B}}$ é uma aplicação linear

$$[f(U)]_{\mathscr{B}} = \sum_{i=0}^{n} c_i [U^i]_{\mathscr{B}}.$$

Ora, do fato adicional

$$[T_1 T_2]_{\mathscr{B}} = [T_1]_{\mathscr{B}} [T_2]_{\mathscr{B}}$$

para quaisquer T_1, T_2 em $L(V, V)$ decorre que

$$[U^i]_{\mathscr{B}} = ([U]_{\mathscr{B}})^i, \qquad 2 \leq i \leq n.$$

Como esta relação também é válida para $i = 0, 1$ obtemos o resultado

(4-17) $$[f(U)]_{\mathscr{B}} = f([U]_{\mathscr{B}}).$$

Em palavras, se U é um operador linear sobre V, a matriz de um polinômio em U, em relação a uma dada base, é o mesmo polinômio na matriz de U.

Teorema 3. *Se* F *é um corpo contendo um número infinito de elementos distintos, a aplicação* $f \to f^{\sim}$ *é um isomorfismo da álgebra dos polinômios sobre* F *na álgebra das funções polinomiais sobre* F.

Demonstração. Pela definição, a aplicação é sobrejetora e se f, g pertencem a $F[x]$, é evidente que

$$(cf + dg)^{\sim} = cf^{\sim} + dg^{\sim}$$

para todos os escalares c e d. Como já mostramos que $(fg)^{\sim} = f^{\sim}g^{\sim}$ basta mostrar que a aplicação é injetora. Para tanto, é suficiente, pela linearidade, demonstrar que $f^{\sim} = 0$ implica $f = 0$. Suponhamos então que f seja um polinômio de grau menor ou igual a n tal que $f^{\sim} = 0$. Sejam $t_0, t_1, \ldots, t_n$ $n + 1$ elementos arbitrários distintos de F. Como $f^{\sim} = 0$, $f(t_i) = 0$ para $i = 0, 1, \ldots, n$, e é uma conseqüência imediata de (4-14) que $f = 0$.

A partir dos resultados da próxima seção, obteremos uma demonstração totalmente diferente deste teorema.

Exercícios

1. Usar a fórmula de interpolação de Lagrange para determinar um polinômio f com coeficientes reais tal que f tenha grau ≤ 3 e $f(-1) = -6$, $f(0) = 2$, $f(1) = -2$, $f(2) = 6$.

2. Sejam α, β, γ e δ números reais. Perguntamos quando é possível determinar um polinômio f sobre R, de *grau não maior que* 2, tal que $f(-1) = \alpha$, $f(1) = \beta$, $f(3) = \gamma$, e $f(0) = \delta$. Demonstrar que isto é possível se, e somente se,

$$3\alpha + 6\beta - \gamma - 8\delta = 0.$$

3. Seja F o corpo dos números reais,

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

e

$$p = (x - 2)(x - 3)(x - 1).$$

(a) Mostrar que $p(A) = 0$.

(b) Sejam P_1, P_2, P_3 os polinômios de Lagrange para $t_1 = 2$, $t_2 = 3$, $t_3 = 1$. Calcular $E_i = P_i(A)$, $i = 1, 2, 3$.

(c) Mostrar $E_1 + E_2 + E_3 = I$, $E_iE_j = 0$ se $i \neq j$, $E_i^2 = E_i$.

(d) Mostrar que $A = 2E_1 + 3E_2 + E_3$.

4. Seja $p = (x - 2)(x - 3)(x - 1)$ e seja T um operador linear arbitrário sobre R^4 tal que $p(T) = 0$. Sejam P_1, P_2, P_3 os polinômios de Lagrange do Exercício 3 e seja $E_i = P_i(T)$, $i = 1, 2, 3$. Demonstrar que

$$E_1 + E_2 + E_3 = I, \quad E_iE_j = 0 \quad \text{se} \quad i \neq j,$$

$$E_i^2 = E_i, \quad \text{e} \quad T = 2E_1 + 3E_2 + E_3.$$

5. Seja n um inteiro positivo e F um corpo. Suponhamos que A seja uma $n \times n$ matriz sobre F e P seja uma $n \times n$ matriz inversível sobre F. Se f é um polinômio arbitrário sobre F, demonstrar que

$$f(P^{-1}AP) = P^{-1}f(A)P.$$

6. Seja F um corpo. Consideremos certos funcionais lineares particulares sobre F, obtidos pelo "cálculo do valor em t":

$$L(f) = f(t).$$

Tais funcionais não são apenas lineares, mas também têm a propriedade de que $L(fg) = L(f)L(g)$. Demonstrar que, se L é um funcional linear qualquer sobre $F[x]$ tal que

$$L(fg) = L(f)L(g)$$

para todos f e g, então $L = 0$ ou existe um t em F tal que $L(f) = f(t)$ para todos f.

4.4 Ideais de Polinômios

Nesta seção preocupamo-nos com resultados que dependem fundamentalmente da estrutura multiplicadora da álgebra dos polinômios sobre um corpo.

Lema. *Suponhamos que* f *e* d *sejam polinômios não-nulos sobre um corpo* F *tais que* $gr(\mathrm{d}) \leq gr(\mathrm{f})$. *Então existe um polinômio* g *em* F[x] *tal que*

$$\mathrm{f} - \mathrm{dg} = 0 \quad ou \quad gr(\mathrm{f} - \mathrm{dg}) < gr(\mathrm{f})$$

Demonstração. Suponhamos que

$$f = a_m x^m + \sum_{i=0}^{m-1} a_i x^i, \qquad a_m \neq 0$$

e que

$$d = b_n x^n + \sum_{i=0}^{n-1} b_i x^i, \qquad b_n \neq 0.$$

Então $m \geq n$ e

$$f - \frac{a_m}{b_n} x^{m-n} d = 0 \quad \text{ou} \quad \text{gr}\left[f - \left(\frac{a_m}{b_n}\right) x^{m-n} d\right] < \text{gr}(f).$$

Assim, podemos tomar $g = \left(\frac{a_m}{b_n}\right) x^{m-n}$.

Usando este lema podemos mostrar que o processo familiar de divisão de polinômios com coeficientes reais ou complexos é possível sobre um corpo arbitrário.

Teorema 4. *Se* f, d *são polinômios sobre um corpo* F *e* d é *diferente de* 0, *então existem polinômios* q, r *em* F[x] *tais que*

(i) f = dq + r.

(ii) r = 0 *ou* *gr*(r) < *gr*(d).

Os polinômios q, r *que satisfazem* (i) *e* (ii) *são únicos.*

Demonstração. Se f é 0 ou se $\text{gr}(f) < \text{gr}(d)$ podemos tomar $q = 0$ e $r = f$. Caso $f \neq 0$ e $\text{gr}(f) \geq \text{gr}(d)$, o lema anterior mostra que podemos escolher um polinômio g tal que $f - dg = 0$ ou $\text{gr}(f - dg) < \text{gr}(f)$. Se $f - dg \neq 0$ e $\text{gr}(f - dg) \geq \text{gr}(d)$ podemos escolher um polinômio h tal que $(f - dg) - dh = 0$ ou

$$\text{gr}[f - d(g + h)] < \text{gr}(f - dg).$$

Continuando este processo enquanto for necessário, obteremos no final polinômios q, r tais que $r = 0$ ou $\text{gr}(r) < \text{gr}(d)$ e $f = dq + r$. Suponhamos agora que também tenhamos $f = dq_1 + r_1$ onde $r_1 = 0$ ou $\text{gr}(r_1) < \text{gr}(d)$. Então $dq + r = dq_1 + r_1$, e $d(q - q_1) = r_1 - r$. Se $q - q_1 \neq 0$ então $d(q - q_1) \neq 0$ e

$$\text{gr}(d) + \text{gr}(q - q_1) = \text{gr}(r_1 - r).$$

Mas como o grau de $r_1 - r$ é menor que o grau de d, isto é impossível e $q - q_1 = 0$. Logo $r_1 - r = 0$.

Definição. *Seja* d *um polinômio não-nulo sobre o corpo* F. *Se* f *está em* F[x], *o teorema anterior mostra que existe no máximo um polinômio* q *em* F[x] *tal que* f = dq. *Se existe um tal* q *dizemos que* d **divide** f, *que* f *é* **divisível** *por* d, *que* f *é um* **múltiplo** *de* d *e denominaremos* q *o* **quociente** *de* f *por* d. *Escreveremos também* q = f/d.

Corolário 1. *Seja* f *um polinômio sobre o corpo* F *e seja* c *um elemento de* F. *Então* f *é divisível por* x − c *se, e somente se,* f(c) = 0.

Demonstração. Pelo teorema, $f = (x - c)q + r$ onde r é um polinômio constante. Pelo Teorema 2,

$$f(c) = 0q(c) + r(c) = r(c).$$

Logo $r = 0$ se, e somente se, $f(c) = 0$.

Definição. *Seja* F *um corpo. Diz-se que um elemento* c *em* F *é uma* **raiz** *ou um* **zero** *de um dado polinômio* f *sobre* F *se* f(c) = 0.

Corolário 2. *Um polinômio* f *de grau* n *sobre um corpo* F *tem no máximo* n *raízes em* F.

Demonstração. O resultado é obviamente verdadeiro para polinômios de grau 0 e grau 1. Suponhamos que seja verdadeiro para polinômios de grau $n - 1$. Se a é uma raiz de f, $f = (x - a)q$ onde q tem grau $n - 1$. Como $f(b) - 0$ se, e somente se, $a = b$ $q(b) = 0$, decorre de nossa hipótese de indução que f tem no máximo n raízes.

O leitor deve notar que o passo principal na demonstração do Teorema 3 é uma conseqüência imediata deste corolário.

As derivadas formais de um polinômio são úteis na discussão de raízes múltiplas. A **derivada** do polinômio

$$f = c_0 + c_1 x + \ldots + c_n x^n$$

é o polinômio

$$f' = c_1 + 2c_2 x + \ldots + nc_n x^{n-1}$$

Usaremos também a notação $Df = f'$. Derivação é linear, isto é, D é um operador linear sobre $F[x]$. Existem as derivadas formais de maior ordem $f'' = D^2 f$, $f^{(3)} = D^3 f$, e assim por diante.

Teorema 5. (Fórmula de Taylor) *Sejam* F *um corpo de característica zero,* c *um elemento de* F *e* n *um inteiro positivo. Se* f *é um polinômio sobre* F *com* $gr(\mathrm{f}) \leq \mathrm{n}$, *então*

$$f = \sum_{k=0}^{n} \frac{(D^k f)}{k!}(c)(x-c)^k.$$

Demonstração. A fórmula de Taylor é uma conseqüência do teorema do binômio e da linearidade dos operadores $D, D^2, \ldots, D^n$. O teorema do binômio é facilmente demonstrado por indução e afirma que

$$(a+b)^m = \sum_{k=0}^{m} \binom{m}{k} a^{m-k} b^k$$

onde

$$\binom{m}{k} = \frac{m!}{k!(m-k)!} = \frac{m(m-1)\ldots(m-k+1)}{1 \cdot 2 \ldots k}$$

é o conhecido coeficiente binomial, dando o número de combinações de m objetos k a k. Pelo teorema do binômio

$$\begin{aligned} x^m &= [c + (x-c)]^m \\ &= \sum_{k=0}^{m} \begin{bmatrix} m \\ k \end{bmatrix} c^{m-k}(x-c)^k \\ &= c^m + mc^{m-1}(x-c) + \ldots + (x-c)^m \end{aligned}$$

e este é o enunciado da fórmula de Taylor para o caso $f = x^m$. Se

$$f = \sum_{m=0}^{n} a_m x^m$$

então $D^k f(c) = \sum_m a_m (D^k x^m)(c)$

e

$$\begin{aligned} \sum_{k=0}^{n} \frac{D^k f(c)}{k!}(x-c)^k &= \sum_k \sum_m a_m \frac{(D^k x^m)}{k!}(c)(x-c)^k \\ &= \sum_m a_m \sum_k \frac{(D^k x^m)}{k!}(c)(x-c)^k \end{aligned}$$

$$= \sum_m a_m x^m$$

$$= f.$$

Observando-se que os polinômios 1, $(x-c), \ldots, (x-c)^n$ são linearmente independentes (ver Exercício 6, Seção 4.2), a Fórmula de Taylor fornece o único método de escrever f como uma combinação linear dos polinômios $(x-c)^k$ $(0 \leq k \leq n)$.

Apesar de não entrarmos em detalhes, talvez seja bom mencionar aqui que a fórmula de Taylor, com uma interpretação apropriada, é também válida para polinômios sobre corpos de característica finita. Se o corpo F tiver característica finita (a soma de um certo número finito de 1's em F é 0) então podemos ter $k! = 0$ em F, caso em que a divisão de $(D^k f)(c)$ por $k!$ não teria sentido. Apesar disso, a divisão de $D^k f$ por $k!$ tem interpretação, pois cada coeficiente de $D^k f$ é um elemento de F multiplicado por um inteiro divisível por $k!$ Se tudo isto parecer muito confuso, aconselhamos ao leitor restringir sua atenção a corpos de característica 0 ou subcorpos do corpo dos números complexos.

Se c for uma raiz do polinômio f, a **multiplicidade** de c como uma raiz de f é o maior inteiro positivo r tal que $(x-c)^r$ divida f.

Obviamente a multiplicidade de uma raiz é menor ou igual ao grau de f. Para polinômios sobre corpos de característica zero, a multiplicidade de c como uma raiz de f, está relacionada com o número de derivadas de f que se anulam em c.

Teorema 6. *Seja* F *um corpo de característica zero e* f *um polinômio sobre* F *com* $gr(\mathrm{f}) \leq \mathrm{n}$. *Então o escalar* c *é uma raiz de* f *de multiplicidade* r *se, e somente se,*

$$(\mathrm{D^k f})(\mathrm{c}) = 0, \qquad 0 \leq \mathrm{k} \leq \mathrm{r}-1$$

$$(\mathrm{D^r f})(\mathrm{c}) \neq 0.$$

Demonstração. Suponhamos que r seja a multiplicidade de c como uma raiz de f. Então existe um polinômio g tal que $f = (x-c)^r g$ e $g(c) \neq 0$. Caso contrário, pelo Corolário 1 do Teorema 4, f seria divisível por $(x-c)^{r+1}$. Pela Fórmula de Taylor aplicada a g

$$f = (x - c)^r \left[\sum_{m=0}^{n-r} \frac{(D^m g)}{m!}(c)\,(x - c)^m \right]$$

$$= \sum_{m=0}^{n-r} \frac{(D^m g)}{m!}(x - c)^{r+m}$$

Como só existe uma maneira de escrever f como combinação linear das potências $(x - c)^k$ $(0 \leq k \leq n)$, segue-se que

$$\frac{(D^k f)(c)}{k!} = \begin{cases} 0 & \text{se} \quad 0 \leq k \leq r - 1 \\ \dfrac{D^{k-r} g(c)}{(k - r)!} & \text{se} \quad r \leq k \leq n \end{cases}$$

Portanto, $D^k f(c) = 0$ para $0 \leq k \leq r - 1$ e $D^r f(c) = g(c) \neq 0$. Reciprocamente, se estas condições estiverem satisfeitas, decorre imediatamente da fórmula de Taylor, que existe um polinômio g tal $f = (x - c)^r g$ e $g(c) \neq 0$. Suponhamos agora que r não seja o maior inteiro positivo tal que $(x - c)^r$ divida f. Então existe um polinômio h tal que $f = (x - c)^{r+1} h$. Mas isto implica, pelo Corolário 2 do Teorema 1, que $g = (x - c)h$; logo $g(c) = 0$, uma contradição.

Definição. *Seja* F *um corpo. Um* **ideal** *em* F[x] *é um subespaço* M *de* F[x] *tal que* fg *pertence a* M *para todo* f *em* F[x] *e todo* g *em* M.

Exemplo 5. Se F é um corpo e d um polinômio sobre F, o conjunto $M = dF[x]$, de todos os múltiplos df de d com f arbitrário em $F[x]$, é um ideal. De fato, M é não-vazio, pois M contém d. Se f, g pertencem a $F[x]$ e c é um escalar, então

$$c(df) - dg = d(cf - g)$$

pertence a M, portanto M é um subespaço. Finalmente, M também contém $(df)g = d(fg)$. O ideal M é denominado o **ideal principal gerado por** d.

Exemplo 6. Sejam $d_1, \ldots, d_n$ um número finito de polinômios sobre F. Então a soma M dos subespaços $d_i F[x]$ é um subespaço e também é um ideal. De fato, suponhamos que p pertença a M. Então existem polinômios $f_1, \ldots, f_n$ em $F[x]$ tais que $p = d_1 f_1 + \ldots + d_n f_n$. Se g é um polinômio arbitrário sobre F, então

$$pg = d_1(f_1 g) + \ldots + d_n(f_n g)$$

de modo que pg também pertence a M. Assim, M é um ideal e dizemos que M é o ideal **gerado** pelos polinômios $d_1, \ldots, d_n$.

Exemplo 7. Seja F um subcorpo do corpo dos números complexos e consideremos o ideal

$$M = (x + 2)F[x] + (x^2 + 8x + 16)F[x].$$

Afirmamos que $M = F[x]$. De fato, M contém

$$x^2 + 8x + 16 - x(x + 2) = 6x + 16$$

logo M contém $6x + 16 - 6(x + 2) = 4$. Assim, o polinômio constante 1 pertence a M, bem como todos os seus múltiplos.

Teorema 7. *Se* F *é um corpo e* M *é um ideal não-nulo arbitrário em* F[x], *existe um único polinômio unitário* d *em* F[x] *tal que* M *seja o ideal principal gerado por* d.

Demonstração. Por hipótese, M contém um polinômio não-nulo; entre todos os polinômios não-nulos em M existe um polinômio d de grau mínimo. Podemos supor d unitário, pois caso contrário podemos multiplicar d por um escalar de modo a torná-lo unitário. Ora, se f pertence a M, o Teorema 4 mostra que $f = dq + r$ onde $r = 0$ ou $\text{gr}(r) < \text{gr}(d)$. Como d está em M, dq e $f - dq = r$ também pertencem a M. Como d é um elemento de M de grau mínimo não podemos ter $\text{gr}(r) < \text{gr}(d)$, portanto $r = 0$. Assim, $M = dF[x]$. Se g é um outro polinômio unitário tal que $M = gF[x]$, então existem polinômios não-nulos p,q tais que $d = gp$ e $g = dq$. Assim, $d = dpq$ e

$$\text{gr}(d) = \text{gr}(d) + \text{gr}(p) + \text{gr}(q).$$

Logo, $\text{gr}(p) = \text{gr}(q) = 0$ e como d, g são unitários, $p = q = 1$. Assim, $d = g$.

É interessante observar que na demonstração acima, utilizamos um caso particular de um fato mais geral e bastante útil, a saber, se p é um polinômio não-nulo em um ideal M e se f é um polinômio em M que não é divisível por p, então $f = pq + r$ onde o "resto" r pertence a M, é diferente de 0 e tem grau menor que o de p. Já usamos este fato no Exemplo 7 para mostrar que o polinômio constante 1 é o gerador unitário do ideal lá considerado. Em princípio, é sempre possível determinar o polinômio unitário que gera um

dado ideal não-nulo, pois podemos, em última análise, obter um polinômio no ideal que tenha grau mínimo por meio de um número finito de divisões sucessíveis.

Corolário. *Se* $p_1, \ldots, p_n$ *são polinômios sobre um corpo* F, *não todos nulos, existe um único polinômio unitário* d *em* F[x] *tal que*

(a) d *está no ideal gerado por* $p_1, \ldots, p_n$;

(b) d *divide cada um dos polinômios* p_i.

Todo polinômio que satisfaz (a) *e* (b) *satisfaz, necessariamente,*

(c) d *é divisível por todo polinômio que divide cada um dos polinômios* $p_1, \ldots, p_n$.

Demonstração. Seja d o gerador unitário do ideal

$$p_1F[x] + \ldots + p_nF[x],$$

Todo elemento deste ideal é divisível por d; assim, cada um dos polinômios p_i é divisível por d. Suponhamos agora que f seja um polinômio que divida cada um dos polinômios $p_1, \ldots, p_n$. Então, existem polinômios $g_1, \ldots, g_n$ tais que $p_i = fg_i$, $1 \leq i \leq n$. Além disso, como d está no ideal

$$p_1F[x] + \ldots + p_nF[x],$$

existem polinômios $q_1, \ldots, q_n$ em $F[x]$ tais que

$$d = p_1q_1 + \ldots + p_nq_n.$$

Assim

$$d = f[g_1q_1 + \ldots + g_nq_n].$$

Mostramos que d é um polinômio unitário que satisfaz (a), (b) e (c). Se d' é um polinômio arbitrário que satisfaz (a) e (b) decorre de (a) e da definição de d que d' é um múltiplo escalar de d e satisfaz (c). Finalmente, se d' é um polinômio unitário temos $d' = d$.

Definição. *Se* $p_1, \ldots, p_n$ *são polinômios sobre um corpo* F, *não todos nulos, o gerador unitário* d *do ideal*

$$p_1F[x] + \ldots + p_nF[x]$$

é denominado o **máximo divisor comum** (m.d.c.) *de* $p_1, \ldots, p_n$ *e é indicado por* $(p_1, \ldots, p_n)$. *Esta terminologia é justificada pelo corolário anterior. Dizemos que os polinômios* $p_1, \ldots, p_n$ *são* **relativamente primos** *se seu máximo divisor comum é* 1 *ou, equivalentemente, se o ideal que eles geram coincide com* $F[x]$.

Exemplo 8. Seja C o corpo dos números complexos. Então
(a) m.d.c. $(x+2,\ x^2+8x+16)=1$ (ver Exemplo 7);
(b) m.d.c. $((x-2)^2(x+i),\ (x-2)(x^2+1)) = (x-2)(x+i)$.

De fato, o ideal

$$(x-2)^2(x+i)F[x] + (x-2)(x^2+1)F[x]$$

contém

$$(x-2)^2(x+i) - (x-2)(x^2+1) = (x-2)(x+i)(i-2).$$

Logo contém $(x-2)(x+i)$, que é unitário e divide

$$(x-2)^2(x+i) \quad \text{e} \quad (x-2)(x^2+1).$$

Exemplo 9. Seja F o corpo dos números racionais e em $[Fx]$ seja M o ideal gerador por

$$(x-1)(x+2)^2,\ (x+2)^2(x-3) \quad \text{e} \quad (x-3).$$

Então M contém

$$\tfrac{1}{2}(x+2)^2[(x-1)-(x-3)] = (x+2)^2$$

e como

$$(x+2)^2 = (x-3)(x+7) - 17$$

M contém o polinômio constante 1. Assim, $M = F[x]$ e os polinômios

$$(x-1)(x+2)^2, (x+2)^2(x-3) \quad \text{e} \quad (x-3)$$

são relativamente primos.

Exercícios

1. Seja Q o corpo dos números racionais. Determinar quais dos seguintes subconjuntos de $Q[x]$ são ideais. Quando o conjunto for um ideal determinar seu gerador unitário.

(a) Todos f de grau par;
(b) Todos f de grau ≥ 5;
(c) Todos f tais que $f(0) = 0$;
(d) Todos f tais que $f(2) = f(4) = 0$;
(e) Todos f na imagem do operador linear T definido por

$$T\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=0}^{n} \frac{c_i}{i+1} x^{i+1}.$$

2. Determinar o m. d. c. de cada um dos seguintes pares de polinômios

(a) $2x^5 - x^3 - 3x^2 - 6x + 4,\ x^4 + x^3 - x^2 - 2x - 2$;
(b) $3x^4 + 8x^2 - 3,\ x^3 + 2x^2 + 3x + 6$;
(c) $x^4 - 2x^3 - 2x^2 - 2x - 3,\ x^3 + 6x^2 + 7x + 1$.

3. Seja A uma $n \times n$ matriz sobre um corpo F. Mostrar que o conjunto dos polinômios f em $F[x]$ tais que $f[A] = 0$ é um ideal.

4. Seja F um subcorpo dos números complexos e seja

$$A = \begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix}.$$

Determinar o gerador unitário do ideal dos polinômios f em $F[x]$ tais que $f(A) = 0$.

5. Seja F um corpo. Mostrar que a interseção de um número arbitrário de ideais em $F[x]$ é um ideal.

6. Seja F um corpo. Mostrar que o ideal gerado por um número finito de polinômios $f_1, \ldots, f_n$ em $F[x]$ é a interseção de todos os ideais que contêm $f_1, \ldots, f_n$.

7. Seja K um subcorpo de um corpo F e suponhamos que f, g sejam polinômios em $K[x]$. Seja M_K o ideal gerado por f e g em $K[x]$ e M_F o ideal que eles geram em $F[x]$. Mostrar que M_K e M_F possuem o mesmo gerador unitário.

4.5 A Decomposição de um Polinômio em Fatores Primos

Nesta seção demonstraremos que cada polinômio sobre o corpo F pode ser escrito como um produto de polinômios "primos". Esta fatoração nos fornece um instrumento eficiente para determinar o máximo divisor comum de um número finito de polinômios e, em particular, fornece um meio efetivo para decidir se os polinômios são relativamente primos.

Definição. *Seja* F *um corpo. Diz-se que um polinômio* f *em* F[x] *é* **redutível sobre** F *se existem polinômios* g, h *em* F[x[*de grau* ≥ 1 *tais que* f = gh *e, em caso contrário, diz-se que* f *é* **irredutível sobre** F. *Um polinômio não constante irredutível sobre* F *é denominado um* **polinômio primo sobre** F *e dizemos, às vezes, que é um* **primo em** F[x].

Exemplo 10. O polinômio $x^2 + 1$ é redutível sobre o corpo C dos números complexos, pois

$$x^2 + 1 = (x + i)(x - i)$$

e os polinômios $x + i$, $x - i$ pertencem a $C[x]$. Por outro lado, $x^2 + 1$ é irredutível sobre o corpo R dos números reais, pois se

$$x^2 + 1 = (ax + b)(a'x + b')$$

com a, a', b, b' em R, então

$$aa' = 1,\ ab' + ba' = 0,\ bb' = 1.$$

Estas relações implicam $a^2 + b^2 = 0$, o que é impossivel com números reais a e b, a menos que $a = b = 0$.

Teorema 8. *Sejam* p, f *e* g *polinômios sobre o corpo* F. *Suponhamos que* p *seja um polinômio primo e que* p *divida o produto* fg. *Então* p *divide* f *ou* p *divide* g.

Demonstração. Não há perda de generalidade se supomos que p é um polinômio primo unitário. O fato de que p é primo diz então simplesmente que os únicos divisores unitários de p são 1 e p. Seja d o m.d.c. de f e p. Ou $d = 1$ ou $d = p$, pois d é um polinômio unitário que divide p. Se $d = p$ então p divide f e já terminamos. Portanto, suponhamos que $d = 1$, isto é, suponhamos que f e p sejam relativamente primos. Demonstraremos que p divide g. Como $(f,p) = 1$, existem polinômios f_0 e p_0 tais que $1 = f_0 f + p_0 p$. Multiplicando por g obtemos

$$\begin{aligned} g &= f_0 fg + p_0 pg \\ &= (fg) f_0 + p(p_0 g). \end{aligned}$$

Como p divide fg, divide também $(fg) f_0$ e certamente p divide $p(p_0 g)$. Assim p divide g.

Corolário. *Se* p *é um primo e divide um produto* $f_1, \ldots, f_n$, *então* p *divide um dos polinômios* $f_1, \ldots, f_n$.

Demonstração. A demonstração é por indução. Para $n = 2$, o resultado é simplesmente o enunciado do Teorema 8. Suponhamos que o resultado seja válido para $n = k$ e que p divide o produto $f_1, \ldots, f_{k+1}$ de certos $(k+1)$ polinômios. Como p divide $(f_1, \ldots, f_k) f_{k+1}$, p divide f_{k+1} ou p divide $f_1, \ldots, f_k$. Pela hipótese de indução, se p divide $f_1, \ldots, f_k$, então p divide f_j para algum j, $1 \leq j \leq k$. Portanto vemos que em qualquer caso p deve dividir algum f_j, $1 \leq j \leq k+1$.

Teorema 9. *Se* F *é um corpo, um polinômio não-constante e unitário em* F[x] *pode ser decomposto como um produto de primos unitários em* F[x] *de uma e, a menos da ordem, somente uma maneira.*

Demonstração. Suponhamos que f seja um polinômio não-constante e unitário sobre F. Como polinômios de grau 1 são irredutíveis, nada há a demonstrar se $\operatorname{gr}(f) = 1$. Suponhamos que f tenha grau $n > 1$. Por indução podemos supor o teorema verdadeiro para todos os polinômios não-constantes e unitários de grau menor que n. Se f é irredutível, f já está decomposto como um produto de primos unitários e, em caso contrário, $f = gh$ onde g e h são polinômios não-constantes e unitários de grau menor que n. Assim g e h podem ser decompostos como produtos de primos unitários em $F[x]$, logo f também pode sê-lo. Suponhamos agora que

$$f = p_1 \cdots p_m = q_1, \cdots q_n$$

onde $p_1, \ldots, p_m$ e $q_1, \ldots, q_n$ são primos unitários em $F[x]$. Então p_m divide o produto $q_1, \ldots, q_n$. Pelo corolário acima, p_m divide algum q_i. Como q_i e p_m são ambos primos e unitários isto significa que

$$q_i = p_m. \tag{4-18}$$

De (4-18) vemos que $m = n = 1$ se $m = 1$ ou $n = 1$, pois

$$\operatorname{gr}(f) = \sum_{i=1}^{m} \operatorname{gr}(p_i) = \sum_{j=1}^{n} \operatorname{gr}(q_j).$$

Neste caso nada resta a demonstrar, portanto podemos supor $m > 1$ e $n > 1$. Reordenando os q podemos supor que $p_m = q_n$ e que

$$p_1, \ldots, p_{m-1} p_m = q_1, \ldots, q_{n-1} p_m.$$

Ora, pelo Corolário 2 do Teorema 1 decorre que

$$p_1, \ldots, p_{m-1} = q_1, \ldots, q_{n-1}.$$

Como o polinômio $p_1, \ldots, p_{m-1}$ tem grau menor que n, nossa hipótese de indução se aplica e mostra que a seqüência $q_1, \ldots, q_{n-1}$ é no máximo uma reordenação da seqüência $p_1, \ldots, p_{m-1}$. Isto, junto com (4-18), mostra que a decomposição de f num produto de primos unitários é única a menos da ordem dos fatores.

Na decomposição acima de um polinômio f não-constante e unitário, alguns dos fatores primos e unitários podem repetir-se. Se $p_1, p_2, \ldots, p_r$ são os primos unitários distintos que ocorrem nesta decomposição de f, então

(4-19) $$f = p_1^{n_1} p_2^{n_2}, \ldots, p_r^{n_r},$$

sendo o expoente n_i o número de vezes que o primo p_i ocorre na decomposição. Esta decomposição é certamente única e é denominada a **decomposição primária** de f. Verifica-se facilmente que todo divisor unitário de f é da forma

(4-20) $$p_1^{m_1}\ p_2^{m_2}, \ldots, p_r^{m_r},\ 0 \leq m_i \leq n_i.$$

De (4-20) decorre que o m.d.c. de um número finito de polinômios não-constantes e unitários $f_1, \ldots, f_s$ é obtido combinando todos aqueles primos unitários que ocorrem simultaneamente nas decomposições de $f_1, \ldots, f_s$. O expoente ao qual cada primo deve ser elevado é o maior número natural para o qual a correspondente potência prima é um fator de cada f_i. Se não existem potências primas (não-triviais) que sejam fatores de cada f_i, os polinômios são relativamente primos.

Exemplo 11. Suponhamos que F seja um corpo e sejam a, b e c elementos distintos de F. Então os polinômios $x - a$, $x - b$ e $x - c$ são primos unitários distintos em $F[x]$. Se m, n e s são inteiros positivos, $(x - c)^s$ é o m.d.c. dos polinômios

$$(x - b)^n (x - c)^s \quad \text{e} \quad (x - a)^m (x - c)^s$$

enquanto que os três polinômios

$$(x - b)^n(x - c)^s,\ (x - a)^m(x - c)^s,\ (x - a)^m(x - b)^n$$

são relativamente primos.

Teorema 10. *Seja* f *um polinômio não-constante e unitário sobre o corpo* F *e seja*

$$f = p_1^{n_1} \dots p_k^{n_k}$$

a decomposição de f *em fatores primos. Para cada* j, $1 \leq j \leq k$, *seja*

$$f_j = f/p_j^{n_j} \quad \prod_{i \neq j} p_i^{n_i}$$

Então $f_1, \dots, f_k$ *são relativamente primos.*

Demonstração. Deixamos a demonstração (fácil) deste fato a cargo do leitor. Enunciamos este teorema principalmente porque pretendemos usá-lo posteriormente.

Teorema 11. *Seja* f *um polinômio sobre o corpo* F *com derivada* f'. *Então* f *é um produto de polinômios distintos e irredutíveis sobre* F *se, e somente se,* f *e* f' *forem relativamente primos.*

Demonstração. Suponhamos que na decomposição de f em fatores primos sobre o corpo F algum polinômio primo p (não constante) esteja repetido. Então $f = p^2h$ para algum h em $F[x]$. Portanto

$$f' = p^2h' + 2pp'h$$

e também p é um divisor de f'. Logo f e f' não são relativamente primos.

Suponhamos $f = p_1, \dots, p_k$ onde $p_1, \dots, p_k$ são polinômios distintos, não constantes e irredutíveis sobre F. Seja $f_j = f/p_j$. Então

$$f' = p_1'f_1 + p_2'f_2 + \dots + p_k'f_k.$$

Seja p um polinômio primo que divida f e f'. Então $p = p_i$ para algum i. Ora, p_i divide f_j para $j \neq i$, e como p_i também divide

$$f' = \sum_{j=1}^{k} p_j'f_j$$

vemos que p_i necessariamente divide $p_i'f_i$. Portanto p_i divide f_i ou divide p_i'. Mas p_i não divide f_i pois $p_1, \dots, p_k$ são distintos. Por-

tanto p_i divide p_i'. Isto não é possível, pois o grau de p_i' é uma unidade menor que o grau de p_i. Concluímos que nenhum primo divide f e f', ou seja, f e f' são relativamente primos.

Definição. *O corpo* F *é dito* **algebricamente fechado** *se todo polinômio primo sobre* F *tem grau* 1.

Dizer que F é algebricamente fechado significa que todo polinômio não-constante irredutível e unitário sobre F é da forma $(x - c)$. Já observamos que cada um destes polinômios é irredutível para qualquer F. Conseqüentemente, pode-se dar uma definição equivalente de um corpo algebricamente fechado como um corpo F tal que todo polinômio não-constante f em $F[x]$ possa ser expresso sob a forma

$$f = c(x - c_1)^{n_1} \dots (x - c_k)^{n_k}$$

onde c é um escalar, $c_1, \dots, c_k$ são elementos distintos de F e $n_1, \dots, n_k$ são inteiros positivos. Ainda uma outra formulação é que, se f é um polinômio não-constante sobre F, então existe um elemento c em F tal que $f(c) = 0$.

O corpo R dos números reais não é algebricamente fechado porque o polinômio $(x^2 + 1)$ é irredutível sobre R mas não tem grau 1, ou porque não existe um número real c tal que $c^2 + 1 = 0$. O chamado Teorema Fundamental da Álgebra afirma que o corpo C dos números complexos é algebricamente fechado. Não demonstraremos este teorema, apesar de usarmo-lo um pouco mais adiante neste livro. A demonstração é omitida em parte pelas limitações de tempo e em parte porque a demonstração depende de uma propriedade "não algébrica" do sistema dos números reais. Para uma das demonstrações possíveis o leitor interessado poderá consultar o livro de Schreier e Sperner citado na bibliografia.

O Teorema Fundamental da Álgebra também esclarece quais são as possibilidades para que um polinômio com coeficientes reais seja decomposto em fatores primos. Se f é um polinômio com coeficientes reais e c é uma raiz complexa de f, então o complexo conjugado $\bar{c}$ também é uma raiz de f. Portanto, todas as raízes complexas, que não sejam reais, necessariamente ocorrem aos pares conjugados e o conjunto todo de raízes tem a forma $\{t_1, \dots, t_k, c_1, \bar{c}_1, \dots, c_r, \bar{c}_r\}$ onde $t_1, \dots, t_k$ são reais e $c_1, \dots, c_r$ são números complexos não reais. Portanto f se decompõe em

$$f = c(x - t_1) \dots (x - t_k)\, p_1 \dots p_r$$

onde p_i é o polinômio quadrático

$$p_i = (x - c_i)(x - \bar{c}_i).$$

Estes polinômios p_i têm coeficientes reais. Concluímos que todo polinômio irredutível sobre o corpo dos números reais tem grau 1 ou 2. Todo polinômio sobre R é o produto de certos fatores lineares, obtidos a partir das raízes reais de f, e certos polinômios quadráticos irredutíveis.

Exercícios

1. Seja p um polinômio unitário sobre o corpo F e sejam f e g polinômios sobre F, relativamente primos. Demonstrar que o m. d. c. de pf e pg é p.

2. Admitindo o Teorema Fundamental da Álgebra, demonstrar o que segue. Se f e g são polinômios sobre o corpo dos números complexos, então m. d. c. $(f, g) = 1$ se, e somente se, f e g não possuem nenhuma raiz comum.

3. Seja D o operador derivação sobre o espaço dos polinômios sobre o corpo dos números complexos. Seja f um polinômio unitário sobre o corpo dos números complexos. Demonstrar que

$$f = (x - c_1) \dots (x - c_k)$$

onde $c_1, \dots, c_k$ são números complexos *distintos* se, e somente se, f e Df são relativamente primos. Em outras palavras, f não possui raiz múltipla se, e somente se, f e Df não possuem nenhuma raiz comum. (Admitir o Teorema Fundamental da Álgebra.)

4. Demonstrar a seguinte generalização da fórmula de Taylor. Sejam f, g e h polinômios sobre um subcorpo dos números complexos, com $gr(f) \leq n$. Então

$$f(g) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(h)(g - h)^k.$$

(Aqui $f(g)$ indica "f de g".)

Para os exercícios restantes, precisaremos da definição seguinte: Se f, g e p são polinômios sobre o corpo F com $p \neq 0$, dizemos que f é **congruente a g módulo p** se $(f - g)$ é divisível por p. Se f é congruente a g módulo p, escrevemos

$$f \equiv g \mod p.$$

5. Demonstrar, para todo polinômio não-nulo p, que a congruência módulo p é uma relação de equivalência.

(*a*) Reflexiva: $f \equiv f \mod p$.

(b) Simétrica: se $f \equiv g \mod p$ e $g \equiv f \mod p$.

(c) Transitiva: se $f \equiv g \mod p$ e $g \equiv h \mod p$, então $f \equiv h \mod p$.

6. Suponhamos que $f \equiv g \mod p$ e $f_1 \equiv g_1 \mod p$.

(a) Demonstrar que $f + f_1 \equiv g + g_1 \mod p$;

(b) Demonstrar que $ff_1 \equiv gg_1 \mod p$.

7. Usar o Exercício 6 para demonstrar o que segue. Se f, g, h e p são polinômios sobre o corpo F e $p \neq 0$ e se $f \equiv g \mod p$, então $h(f) \equiv h(g) \mod p$.

8. Se p é um polinômio irredutível e $fg = 0 \mod p$, demonstrar que ou $f \equiv 0 \mod p$ ou $g \equiv 0 \mod p$. Dar um exemplo que mostre que isto é falso se p não é irredutível.

Capítulo 5
DETERMINANTES

5.1 Anéis Comutativos

Neste capítulo demonstraremos os fatos essenciais sobre determinantes de matrizes quadradas. Faremos isto não apenas para matrizes sobre um corpo, mas também para matrizes cujos elementos sejam "escalares" de um tipo mais geral. Existem duas razões para esta generalidade. Primeiro, em certos pontos do próximo capítulo, será necessário tratarmos de determinantes de matrizes cujos elementos são polinômios. Segundo, no tratamento de determinantes que apresentamos, um dos axiomas da definição de um corpo não desempenha nenhum papel, a saber, o axioma que garante a existência de um inverso multiplicativo para todo elemento não-nulo. Por estas razões, é conveniente desenvolver a teoria dos determinantes para matrizes cujos elementos sejam pertencentes a um anel comutativo com elemento unidade.

Definição. *Um* **anel** *é um conjunto* K*, munido de duas operações* $(x, y) \rightarrow x + y$ *e* $(x, y) \rightarrow xy$ *que satisfazem*

(1) K *é um grupo comutativo em relação à operação* $(x, y) \rightarrow x + y$ (K *é um grupo comutativo em relação à adição*);

(2) $(xy)z = x(yz)$ (*a multiplicação é associativa*);

(3) $x(y + z) = xy + xz$; $(y + z)x = yz + zx$ (*valem as duas leis distributivas*).

Se $xy = yx$ *para todos* x *e* y *em* K*, dizemos que o anel* K *é* **comutativo.** *Se existe um elemento* 1 *em* K *tal que* $1x = x1 = x$ *para todo* x*, dizemos que* K *é um* **anel com elemento unidade** *e* 1 *é denominado o* **elemento unidade de** K.

Estamos interessados em anéis comutativos com elemento unidade. Tal anel pode ser descrito rapidamente como um conjunto *K*, munido de duas operações que satisfazem todos os axiomas da

definição de um corpo dado no Capítulo 1, com exceção talvez do axioma (8) e da condição $1 \neq 0$. Assim, um corpo é um anel comutativo com elemento unidade diferente de zero tal que a cada x não-nulo corresponde um elemento x^{-1} tal que $xx^{-1} = 1$. O conjunto dos inteiros, com as operações usuais, é um anel comutativo com elemento unidade mas não é um corpo. Outro anel comutativo com elemento unidade é o conjunto dos polinômios sobre um corpo, com a adição e a multiplicação que definimos para polinômios.

Se K é um anel comutativo com elemento unidade, definimos uma $m \times n$ matriz sobre K como sendo uma função A do conjunto dos pares (i, j) de inteiros, $1 \leq i \leq m$, $1 \leq j \leq n$, em K. Como sempre, representamos tal matriz por uma tabela retangular com m linhas e n colunas. A soma e o produto de matrizes sobre K são definidas como para matrizes sobre um corpo

$$(A + B)_{ij} = A_{ij} + B_{ij}$$
$$(AB)_{ij} = \sum_k A_{ik}B_{kj}$$

sendo que a soma é definida quando A e B têm o mesmo número de linhas e o mesmo número de colunas e o produto é definido quando o número de colunas de A é igual ao número de linhas de B. As propriedades algébricas básicas destas operações são ainda válidas. Por exemplo

$$A(B + C) = AB + AC, \qquad (AB)C = A(BC), \quad \text{etc.}$$

Como no caso de corpos, referir-nos-emos aos elementos de K como escalares. Podemos então definir combinações lineares das linhas ou colunas de uma matriz como fizemos anteriormente. A grosso modo, tudo o que fizemos para matrizes sobre um corpo é válido para matrizes sobre K, excetuando-se os resultados que dependiam da possibilidade de "dividir" em K.

5.2 Funções Determinantes

Seja K um anel comutativo com elemento unidade. Desejamos associar a cada $n \times n$ matriz (quadrada) sobre K um escalar (elemento de K) que será conhecido como o determinante da matriz. É possível definir o determinante de uma matriz quadrada A simplesmente escrevendo uma fórmula para este determinante em

função dos elementos de A. Pode-se então deduzir as diversas propriedades dos determinantes a partir desta fórmula. Contudo, tal fórmula é bastante complicada e a fim de ganhar alguma vantagem técnica, procederemos como segue. Definiremos uma "função determinante" sobre $K^{n \times n}$ como uma função que associa a cada $n \times n$ matriz sobre K um escalar, sendo que essa função possui propriedades especiais. É linear como uma função de cada uma das linhas da matriz; seu valor é 0 sobre qualquer matriz que tenha duas linhas iguais e seu valor sobre a $n \times n$ matriz unidade é 1. Demonstraremos que existe uma tal função e depois que é única, isto é, que existe exatamente uma tal função. Ao demonstrarmos a unicidade será obtida uma fórmula explícita para o determinante, junto com muitas de suas úteis propriedades.

Esta seção será dedicada à definição da "função determinante" e à demonstração de que existe pelo menos uma tal função.

Definição. *Seja* K *um anel comutativo com elemento unidade,* n *um inteiro positivo e seja* D *uma função que associa a cada* n × n *matriz* A *sobre* K *um escalar* D(A) *em* K. *Dizemos que* D *é* **n-linear** *se para cada* i, $1 \leq i \leq n$, D *é uma função linear da* i-*ésima linha quando as outras* (n − 1) *linhas são mantidas fixas.*

Esta definição exige algum esclarecimento. Se D é uma função de $K^{n \times n}$ em K e se $\alpha_1, \ldots, \alpha_n$ são as linhas da matriz A, escrevamos também

$$D(A) = D(\alpha_1, \ldots, \alpha_n)$$

isto é, consideremos D também como uma função das linhas de A. A afirmação de que D é n-linear significa então

$$D(\alpha_1, \ldots, c\alpha_i + \alpha_i', \ldots, \alpha_n) = cD(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) + D(\alpha_1, \ldots, \alpha_i', \ldots, \alpha_n). \tag{5-1}$$

Se fixamos todas as linhas exceto a linha i e consideramos D como uma função da i-ésima linha, em geral é conveniente indicar $D(A)$ por $D(\alpha_i)$. Assim, podemos abreviar (5-1) para

$$D(c\alpha_i + \alpha_i') = cD(\alpha_i) + D(\alpha_i')$$

desde que esteja claro o que isto significa.

Exemplo 1. Sejam $k_1, \ldots, k_n$ inteiros positivos, $1 \leq k_i \leq n$ e seja a um elemento de K. Para cada $n \times n$ matriz A sobre K definamos

$$D(A) = aA(1, k_1) \ldots A(n, k_n). \tag{5-2}$$

Então, a função D definida por (5-2) é n-linear. De fato, considerando D como uma função da i-ésima linha de A, com as outras linhas fixas, podemos escrever

$$D(\alpha_i) = A(i, k_i)b$$

onde b é algum elemento fixo de K. Seja $\alpha'_i = (A'_{i1}, \ldots, A'_{in})$. Então temos

$$\begin{aligned} D(c\alpha_i + \alpha'_i) &= [cA(i, k_i) + A'(i, k_i)]b \\ &= cD(\alpha_i) + D(\alpha'_i). \end{aligned}$$

Assim, D é uma função linear de cada uma das linhas de A.

Uma particular função n-linear deste tipo é

$$D(A) = A_{11}A_{22} \ldots A_{nn}.$$

Em outras palavras, o "produto dos elementos diagonais" é uma função n-linear sobre $K^{n \times n}$.

Exemplo 2. Determinemos todas as funções bilineares (2-lineares) sobre 2×2 matrizes sobre K. Seja D uma dessas funções. Indicando as linhas da 2×2 matriz unidade por ε_1, ε_2, temos

$$D(A) = D(A_{11}\varepsilon_1 + A_{12}\varepsilon_2, A_{21}\varepsilon_1 + A_{22}\varepsilon_2).$$

Usando o fato de que D é bilinear, (5-1) implica

$$\begin{aligned} D(A) &= A_{11}D(\varepsilon_1, A_{21}\varepsilon_1 + A_{22}\varepsilon_2) + A_{12}D(\varepsilon_2, A_{21}\varepsilon_1 + A_{22}\varepsilon_2) \\ &= A_{11}A_{21}D(\varepsilon_1, \varepsilon_1) + A_{11}A_{22}D(\varepsilon_1, \varepsilon_2) \\ &\qquad + A_{12}A_{21}D(\varepsilon_2, \varepsilon_1) + A_{12}A_{22}D(\varepsilon_2, \varepsilon_2). \end{aligned}$$

Assim, D é completamente determinada pelos quatro escalares

$$D(\varepsilon_1, \varepsilon_1), \quad D(\varepsilon_1, \varepsilon_2), \quad D(\varepsilon_2, \varepsilon_1) \quad \text{e} \quad D(\varepsilon_2, \varepsilon_2).$$

O leitor deverá achar fácil verificar o seguinte: se a, b, c, d são quatro escalares arbitrários em K e se definimos

$$D(A) = A_{11}A_{21}a + A_{11}A_{22}b + A_{12}A_{21}c + A_{12}A_{22}d,$$

então D é uma função bilinear sobre as 2×2 matrizes sobre K e

$$\begin{aligned} D(\varepsilon_1, \varepsilon_1) &= a, & D(\varepsilon_1, \varepsilon_2) &= b, \\ D(\varepsilon_2, \varepsilon_1) &= c, & D(\varepsilon_2, \varepsilon_2) &= d. \end{aligned}$$

Lema. *Uma combinação linear de funções* n-*lineares é* n-*linear.*

Demonstração. Basta demonstrar que uma combinação linear de duas funções n-lineares é n-*linear.* Sejam D e E funções n-lineares. Se a e b pertencem a K, a combinação linear $aD + bE$ é evidentemente definida por

$$(aD + bE)(A) = aD(A) + bE(A).$$

Logo, fixando todas as linhas exceto a linha i

$$\begin{aligned} (aD + bE)(c\alpha_i + \alpha_i') &= aD(c\alpha_i + \alpha_i') + bE(c\alpha_i + \alpha_i') \\ &= acD(\alpha_i) + aD(\alpha_i') + bcE(\alpha)_i + bE(\alpha_i') \\ &= c(aD + bE)(\alpha_i) + (aD + bE)(\alpha_i'). \end{aligned}$$

Se K é um corpo e V é o conjunto das $n \times n$ matrizes sobre K, o lema acima diz o seguinte: o conjunto das funções n-lineares sobre V é um subespaço do espaço de todas as funções de V em K.

Exemplo 3. Seja D a função definida sobre as 2×2 matrizes sobre K por

$$D(A) = A_{11}A_{22} - A_{12}A_{21}. \tag{5-3}$$

Ora, D é a soma de duas funções do tipo descrito no Exemplo 1:

$$\begin{aligned} D &= D_1 + D_2 \\ D_1(A) &= A_{11}A_{22} \\ D_2(A) &= -A_{12}A_{21}. \end{aligned}$$

Pelo Lema acima, D é uma função bilinear. O leitor que tenha tido alguma experiência com determinantes não achará este fato surpreendente, pois reconhecerá (5-3) como a definição usual do deter-

minante de uma 2×2 matriz. É claro que a função D que acabamos de definir não é uma função bilinear típica. Ela possui muitas propriedades particulares. Vejamos algumas dessas propriedades. Primeiro, se I é a 2×2 matriz unidade, então $D(I) = 1$, isto é, $D(\varepsilon_1, \varepsilon_2) = 1$. Segundo, se as duas linhas de A são iguais, então

$$D(A) = A_{11}A_{12} - A_{12}A_{11} = 0.$$

Terceiro, se A' é a matriz obtida de uma 2×2 matriz A permutando suas linhas, então $D(A') = -D(A)$; de fato,

$$\begin{aligned} D(A') &= A'_{11}A'_{22} - A'_{12}A'_{21} \\ &= A_{21}A_{12} - A_{22}A_{11} \\ &= -D(A). \end{aligned}$$

Definição. *Seja* D *uma função* n-*linear. Dizemos que* D *é* **alternada** (*ou* **alternante**) *se as duas condições seguintes estão satisfeitas:*

(a) D(A) = 0 *sempre que duas linhas de* A *são iguais.*

(b) *Se* A′ *é uma matriz obtida de* A *permutando duas linhas de* A, *então* D(A′) = − D(A).

Demonstraremos abaixo que toda função n-linear D que satisfaz (a), automaticamente satisfaz (b). Colocamos as duas propriedades na definição de função n-linear alternada por conveniência. O leitor provavelmente notará que se D satisfaz (b) e A é uma matriz com duas linhas iguais, então $D(A) = -D(A)$. Somos tentados a concluir que D também satisfaz a condição (a). Isto é verdade, por exemplo, se K é um corpo no qual $1 + 1 \neq 0$, mas em geral (a) não é uma conseqüência de (b).

Definição. *Seja* K *um anel comutativo com elemento unidade e seja* n *um inteiro positivo. Suponhamos que* D *seja uma função das* n × n *matrizes sobre* K *em* K. *Dizemos que* D *é uma* **função determinante** *se* D *é* n-*linear, alternada e* D(I) = 1.

Como afirmamos anteriormente, mostraremos no final que existe exatamente uma função determinante sobre $n \times n$ matrizes sobre K. Isto pode ser visto facilmente para 1×1 matrizes $A = [a]$ sobre K. A função D dada por $D(A) = a$ é uma função determinante e é evidentemente a única função determinante sobre 1×1 matrizes. Estamos em condições de eliminar também o caso $n = 2$. Demonstramos, no Exemplo 3, que a função

$$D(A) = A_{11}A_{22} - A_{12}A_{21}$$

é uma função determinante. Além disso, a fórmula apresentada no Exemplo 2 mostra que D é a única função determinante sobre 2×2 matrizes, pois demonstramos que para qualquer função bilinear D

$$D(A) = A_{11}A_{21}D(\varepsilon_1, \varepsilon_1) + A_{11}A_{22}D(\varepsilon_1, \varepsilon_2) + A_{12}A_{21}D(\varepsilon_2, \varepsilon_1) + A_{12}A_{22}D(\varepsilon_2, \varepsilon_2).$$

Se D é alternada, então

$$D(\varepsilon_1, \varepsilon_1) = D(\varepsilon_2, \varepsilon_2) = 0$$

e

$$D(\varepsilon_2, \varepsilon_1) = -D(\varepsilon_1, \varepsilon_2) = -D(I).$$

Se D satisfaz também $D(I) = 1$, então

$$D(A) = A_{11}A_{22} - A_{12}A_{21}.$$

Exemplo. Seja F um corpo e seja D uma função trilinear alternada, arbitrária, sobre as 3×3 matrizes sobre o anel de polinômios $F[x]$.

Seja

$$A = \begin{bmatrix} x & 0 & -x^2 \\ 0 & 1 & 0 \\ 1 & 0 & x^3 \end{bmatrix}$$

Se indicarmos as linhas da 3×3 matriz unidade por $\varepsilon_1, \varepsilon_2, \varepsilon_3$, então

$$D(A) = D(x\varepsilon_1 - x^2\varepsilon_3, \varepsilon_2, \varepsilon_1 + x^3\varepsilon_3).$$

Como D é uma função linear de cada linha

$$\begin{aligned} D(A) &= xD(\varepsilon_1, \varepsilon_2, \varepsilon_1 + x^3\varepsilon_3) - x^2D(\varepsilon_3, \varepsilon_2, \varepsilon_1 + x^3\varepsilon_3) \\ &= xD(\varepsilon_1,\varepsilon_2,\varepsilon_1) + x^4D(\varepsilon_1,\varepsilon_2,\varepsilon_3) - x^2D(\varepsilon_3,\varepsilon_2,\varepsilon_1) - x^5D(\varepsilon_3,\varepsilon_2,\varepsilon_3). \end{aligned}$$

Como D é alternada decorre que

$$D(A) = (x^4 + x^2)\,D(\varepsilon_1, \varepsilon_2, \varepsilon_3).$$

Lema. *Seja* D *uma função bilinear com a propriedade de que* $D(A) = 0$ *para todas* 2×2 *matrizes* A *sobre* K *que tenham linhas iguais. Então* D *é alternada.*

Demonstração. O que precisamos mostrar é que se A é uma 2×2 matriz e A' é obtida transpondo-se as linhas de A, então $D(A') = -D(A)$. Se as linhas de A são α e β, isto significa que precisamos mostrar que $D(\beta, \alpha) = -D(\alpha, \beta)$. Como D é bilinear

$$D(\alpha + \beta, \alpha + \beta) = D(\alpha, \alpha) + D(\alpha, \beta) + D(\beta, \alpha) + D(\beta, \beta).$$

Por nossa hipótese, $D(\alpha + \beta, \ \alpha + \beta) = D(\alpha, \alpha) = D(\beta, \beta) = 0$. Portanto,

$$0 = D(\alpha, \beta) + D(\beta, \alpha).$$

Lema. *Seja* D *uma função* n-*linear sobre* $n \times n$ *matrizes sobre* K. *Suponhamos que* D *tenha a propriedade de que* $D(A) = 0$ *sempre que duas linhas adjacentes de* A *sejam iguais. Então* D *é alternada.*

Demonstração. Precisamos mostrar que $D(A) = 0$ quando duas quaisquer linhas de A são iguais e que $D(A') = -D(A)$ e A' é obtida transpondo-se (trocando entre si) duas quaisquer linhas de A. Em primeiro lugar suponhamos que A' seja obtida pela transposição de duas linhas adjacentes de A. O leitor deve verificar que o argumento usado na demonstração do lema anterior se aplica ao presente caso e nos fornece $D(A') = -D(A)$.

Seja agora B obtida transpondo-se as linhas i e j de A, onde $i < j$. Podemos obter B a partir de A por uma sucessão de transposições de pares de linhas adjacentes. Começamos transpondo a linha i com a linha $(i + 1)$ e continuamos até que as linhas estejam na ordem

$$\alpha_1, \ldots, \alpha_{i-1}, \ \alpha_{i+1}, \ldots, \alpha_j, \ \alpha_i, \alpha_{j+1}, \ldots, \alpha_n.$$

Isto requer $k = j - i$ transposições de linhas adjacentes. Agora, movemos α_j para a i-ésima posição usando $(k - 1)$ transposições de linhas adjacentes. Desta maneira obtivemos B a partir de A por meio de $k + (k - 1) = 2k - 1$ transposições de linhas adjacentes. Assim,

$$D(B) = (-1)^{2k-1} D(A) = -D(A).$$

Suponhamos que A seja uma $n \times n$ matriz arbitrária com duas linhas iguais, digamos $\alpha_i = \alpha_j$, com $i < j$. Se $j = i + 1$, então A tem duas linhas iguais e adjacentes e $D(A) = 0$. Se $j > i + 1$ transpomos α_{i+1} e α_j e a matriz resultante B possui duas linhas iguais e adjacentes, portanto $D(B) = 0$. Por outro lado $D(B) = -D(A)$, portanto $D(A) = 0$.

Definição. *Se* n > 1 *e* A *é uma* n × n *matriz sobre* K, *indiquemos por* A(i|j) *a* (n − 1) × (n − 1) *matriz obtida de* A *retirando-se a* i-*ésima linha e a* j-*ésima coluna. Se* D *é uma função* (n − 1)-*linear e* A *é uma* n × n *matriz, colocamos* $D_{ij}(A) = D[A(i|j)]$.

Teorema 1. *Seja* n > 1 *e seja* D *uma função* (n − 1)-*linear alternada sobre as* (n − 1) × (n − 1) *matrizes sobre* K. *Para cada* j, 1 ≤ j ≤ n, *a função* E_j *definida por*

$$E_j(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} D_{ij}(A) \tag{5-4}$$

é uma função n-*linear alternada sobre* n × n *matrizes* A. *Se* D *é uma função determinante, cada* E_j *também o é.*

Demonstração. Se A é uma $n \times n$ matriz, $D_{ij}(A)$ é independente da i-ésima linha de A. Como D é $(n-1)$-linear, é claro que D_{ij} é uma função linear de qualquer linha exceto a linha i. Portanto $A_{ij}D_{ij}(A)$ é uma função n-linear de A. Uma combinação linear de funções n-lineares é n-linear, logo E_j é n-linear. Para demonstrar que E_j é alternada, bastará demonstrar que $E_j(A) = 0$ sempre que A tiver duas linhas iguais e adjacentes. Suponhamos que $\alpha_k = \alpha_{k+1}$. Se $i \neq k$ e $i \neq k+1$, a matriz $(A(i|j)$ tem duas linhas iguais e então $D_{ij}(A) = 0$. Portanto,

$$E_j(A) = (-1)^{k+j} A_{kj} D_{kj}(A) + (-1)^{k+1+j} A_{(k+1)j} D_{(k+1)j}(A).$$

Como $\alpha_k = \alpha_{k+1}$,

$$A_{kj} = A_{(k+1)j} \quad \text{e} \quad A(k|j) = A(k+1|j).$$

Então é evidente que $E_j(A) = 0$.

Suponhamos agora que D seja uma função determinante. Se $I^{(n)}$ é a $n \times n$ matriz unidade, então $I^{(n)}(j/j)$ é a $(n-1) \times (n-1)$ matriz unidade $I^{(n-1)}$. Como $I^{(n)}_{ij} = \delta_{ij}$, decorre de (5-4) que

$$E_j(I^{(n)}) = D(I^{(n-1)}). \tag{5-5}$$

Ora, $D(I^{(n-1)}) = 1$, de modo que $E_j(I^{(n)}) = 1$ e E_j é uma função determinante.

Corolário. *Seja* K *um anel comutativo com elemento unidade e seja* n *um inteiro positivo. Existe pelo menos uma função determinante sobre* $K^{n \times n}$.

Demonstração. Demonstramos a existência de uma função determinante sobre 1×1 matrizes sobre K e mesmo sobre 2×2 matrizes sobre K. O Teorema 1 nos diz explicitamente como construir uma função determinante sobre $n \times n$ matrizes, a partir de uma função determinante sobre $(n-1) \times (n-1)$ matrizes. O corolário decorre por indução.

Exemplo 5. Se B é uma 2×2 matriz sobre K façamos

$$|B| = B_{11}B_{22} - B_{12}B_{21}.$$

Então $|B| = D(B)$, onde D é a função determinante sobre 2×2 matrizes. Mostramos que esta função sobre $K^{2\times 2}$ é única. Seja

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{32} & A_{32} & A_{33} \end{bmatrix}$$

uma 3×3 matriz sobre K. Definindo E_1, E_2, E_3 como em (5-4), então

$$\text{(5-6)} \quad E_1(A) = A_{11}\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{21}\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} + A_{31}\begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix}$$

$$\text{(5-7)} \quad E_2(A) = -A_{12}\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{22}\begin{vmatrix} A_{11} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} - A_{32}\begin{vmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{vmatrix}$$

$$\text{(5-8)} \quad E_3(A) = A_{13}\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix} - A_{23}\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix} + A_{33}\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}$$

Decorre do Teorema 1 que E_1, E_2, E_3 são funções determinantes. Na realidade, como mostraremos posteriormente, $E_1 = E_2 = E_3$, mas isto ainda não é evidente mesmo neste caso simples. Poder-se-ia, contudo, verificar diretamente este fato desenvolvendo-se cada uma das expressões acima. Em vez de fazer isto, daremos exemplos particulares.

(a) Seja $K = R[x]$ e

$$A = \begin{bmatrix} x-1 & x^2 & x^3 \\ 0 & x-2 & 1 \\ 0 & 0 & x-3 \end{bmatrix}.$$

Então

$$E_1(A) = (x-1)\begin{vmatrix} x-2 & 1 \\ 0 & x-3 \end{vmatrix} = (x-1)(x-2)(x-3)$$

$$E_2(A) = -x^2\begin{vmatrix} 0 & 1 \\ 0 & x-3 \end{vmatrix} + (x-2)\begin{vmatrix} x-1 & x^3 \\ 0 & x-3 \end{vmatrix}$$

$$= (x-1)(x-2)(x-3)$$

e

$$E_3(A) = x^3\begin{vmatrix} 0 & x-2 \\ 0 & 0 \end{vmatrix} - \begin{vmatrix} x-1 & x^2 \\ 0 & 0 \end{vmatrix} + (x-3)\begin{vmatrix} x-1 & x^2 \\ 0 & x-2 \end{vmatrix}$$

$$= (x-1)(x-2)(x-3).$$

(b) Seja $K = R$ e

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Então

$$E_1(A) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

$$E_2(A) = -\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 1$$

$$E_3(A) = -\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 1 \cdot$$

Exercícios

1. Cada uma das expressões seguintes, define uma função D sobre o conjunto das 3×3 matrizes sobre o corpo dos números reais. Em quais destes casos D é uma função trilinear?

(a) $D(A) = A_{11} + A_{22} + A_{33}$;
(b) $D(A) = (A_{11})^2 + 3A_{11}A_{22}$;
(c) $D(A) = A_{11}A_{12}A_{33}$;
(e) $D(A) = 0$;
(f) $D(A) = 1$.

2. Verificar diretamente que as três funções E_1, E_2, E_3 definidas por (5-6), (5-7) e (5-8) são idênticas.

3. Seja K um anel comutativo com elemento unidade. Se A é uma 2×2 matriz sobre K, a *adjunta clássica* de A é a 2×2 matriz adj A definida por

$$\text{adj } A = \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}.$$

Se det indica a única função determinante sobre 2×2 matrizes sobre K, mostrar que

(a) $(\text{adj } A)A = A(\text{adj } A) = (\det A)I$;

(b) $\det(\text{adj } A) = \det(A)$;

(c) $\text{adj}(A^t) = (\text{adj } A)^t$.

(A^t indica a transposta de A.)

4. Seja A uma 2×2 matriz sobre um corpo F. Mostrar que A é inversível se, e somente se, $\det A \neq 0$. Quando A é inversível, dar uma fórmula para A^{-1}.

5. Seja A uma 2×2 matriz sobre um corpo F e suponhamos que $A^2 = 0$. Mostrar para cada escalar c que $\det(cI - A) = c^2$.

6. Seja K um subcorpo do corpo dos números complexos e n um inteiro positivo. Sejam $j_1, \ldots, j_n$ e $k_1, \ldots, k_n$ inteiros positivos menores ou iguais a n. Para uma $n \times n$ matriz A sobre K definamos

$$D(A) = A(j_1, k_1)\, A(j_2, k_2) \ldots A(j_n, k_n).$$

Demonstrar que D é n-linear se, e somente se, os inteiros $j_1, \ldots, j_n$ são distintos.

7. Seja K um anel comutativo com elemento unidade. Mostrar que a função determinante sobre as 2×2 matrizes A sobre K é alternada e bilinear como uma função das colunas de A.

8. Seja K um anel comutativo com elemento unidade. Definamos uma função D sobre as 3×3 matrizes sobre K pela regra

$$D(A) = A_{11} \det\begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} - A_{12} \det\begin{bmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{bmatrix} + A_{13} \det\begin{bmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}.$$

Mostrar que D é alternada e trilinear como uma função das colunas de A.

9. Seja K um anel comutativo com elemento unidade e D uma função n-linear alternada sobre as $n \times n$ matrizes sobre K. Mostrar que

(a) $D(A) = 0$, se uma das linhas de A é 0.

(b) $D(B) = D(A)$ se B é obtida a partir de A somando-se um múltiplo escalar de uma linha de A a outra.

10. Seja F um corpo, A uma 2×3 matriz sobre F e (c_1, c_2, c_3) o vetor em F^3 definido por

$$c_1 = \begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix} \quad c_2 = \begin{vmatrix} A_{13} & A_{11} \\ A_{23} & A_{21} \end{vmatrix} \quad c_3 = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}.$$

Mostrar que

(a) posto $(A) = 2$ se, e somente se, $(c_1, c_2, c_3) \neq 0$;

(b) se A tem posto 2, então (c_1, c_2, c_3) é uma base do espaço-solução do sistema de equações $AX = 0$.

11. Seja K um anel comutativo com elemento unidade e seja D uma função bilinear alternada sobre as 2×2 matrizes sobre K. Mostrar que $D(A) = (\det A)\, D(I)$ para qualquer A. Utilizar agora este resultado (não são permitidos cálculos com os elementos) para mostrar que $\det(AB) = \det(A) \det(B)$ para quaisquer 2×2 matrizes A e B sobre K.

12. Seja F um corpo e D uma função sobre as $n \times n$ matrizes sobre F (com valores em F). Suponhamos que $D(AB) = D(A)\, D(B)$ para quaisquer A, B. Mostrar que ou $D(A) = 0$ para qualquer A ou $D(I) = 1$. No último caso mostrar que $D(A) \neq 0$ sempre que A é inversível.

13. Seja R o corpo dos números reais e seja D uma função sobre as 2×2 matrizes sobre R, com valores em R, tal que $D(AB) = D(A)\, D(B)$ para quaisquer A, B. Suponhamos também que

$$D\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) \neq D\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

Demonstrar o seguinte:

(a) $D(0) = 0$;

(b) $D(A) = 0$ se $A^2 = 0$;

(c) $D(B) = -D(A)$ se B é obtida transpondo-se duas linhas (ou colunas de A);

(d) $D(A) = 0$ se uma linha (ou coluna) de A é 0;

(e) $D(A) = 0$ sempre que A é singular.

14. Seja A uma 2×2 matriz sobre um corpo F. Então o conjunto das matrizes da forma $f(A)$, onde f é um polinômio sobre F, é um anel comutativo K com elemento unidade. Se B é uma 2×2 matriz sobre K, o determinante de B é então uma 2×2 matriz sobre F, da forma $f(A)$. Suponhamos que I seja a 2×2 matriz unidade sobre F e que B seja a 2×2 matriz sobre K

$$B = \begin{bmatrix} A - A_{11}I & -A_{12}I \\ -A_{21}I & A - A_{22}I \end{bmatrix}.$$

Mostrar que $\det B = f(A)$, onde $f = x^2 - (A_{11} + A_{22})x + \det A$, e também que $f(A) = 0$.

5.3 Permutações e a Unicidade dos Determinantes

Nesta seção demonstraremos a unicidade da função determinante sobre as $n \times n$ matrizes sobre K. A demonstração nos levará,

de modo bem natural, a considerar permutações e algumas de suas propriedades básicas.

Suponhamos que D seja uma função n-linear alternada sobre as $n \times n$ matrizes sobre K. Seja A uma $n \times n$ matriz sobre K com linhas $\alpha_1, \alpha_2, \ldots, \alpha_n$. Se indicarmos as linhas da $n \times n$ matriz unidade sobre K por $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, então

(5-9)
$$\alpha_i = \sum_{j=1}^{n} A(i, j)\varepsilon_j, \quad 1 \leq i \leq n.$$

Logo

$$\begin{aligned} D(A) &= D\Big(\sum_j A(1,j)\varepsilon_j, \alpha_2, \ldots, \alpha_n\Big) \\ &= \sum_j A(1, j)\, D(\varepsilon_j, \alpha_2, \ldots, \alpha_n). \end{aligned}$$

Se agora substituirmos α_2 por $\sum_k A(2, k)\varepsilon_k$ veremos que

$$D(\varepsilon_j, \alpha_2, \ldots, \alpha_n) = \sum_k A(2, k)\, D(\varepsilon_j, \varepsilon_k, \ldots, \alpha_n).$$

Assim

$$D(A) = \sum_{j,k} A(1,j)\, A(2, k)\, D(\varepsilon_j, \varepsilon_k, \ldots, \alpha_n).$$

Em $D(\varepsilon_j, \varepsilon_k, \ldots, \alpha_n)$ substituímos em seguida α_3 por $\Sigma\, A(3, l)\varepsilon_l$ e assim por diante. Obtemos finalmente uma expressão complicada mas teoricamente importante para $D(A)$, a saber

(5-10)
$$D(A) = \sum_{k_1, k_2, \ldots, k_n} A(1, k_1) A(2, k_2) \ldots A(n, k_n)\, D(\varepsilon_{k_1}, \varepsilon_{k_2}, \ldots, \varepsilon_{k_n}).$$

Em (5-10) a soma é estendida a todas as seqüências $(k_1, k_2, \ldots, k_n)$ de inteiros positivos menores ou iguais a n. Isto mostra que D é uma soma finita de funções do tipo descrito por (5-2). Deve-se notar que (5-10) é uma conseqüência apenas da hipótese de que D é n-linear e também que um caso particular de (5-10) foi obtido no Exemplo 2. Como D é alternada,

$$D = (\varepsilon_{k_1}, \varepsilon_{k_2}, \ldots, \varepsilon_{k_n}) = 0$$

sempre que dois dos índices k_i são iguais. Uma seqüência $(k_1, k_2, \ldots, \ldots, k_n)$ de inteiros positivos menores ou iguais a n, com a proprie-

dade de não existirem dois k_i iguais, é denominada uma **permutação de grau** n. Portanto, em (5-10), precisamos somar considerando apenas as seqüências que sejam permutações de grau n.

Como uma seqüência finita, ou n-upla, é uma função definida sobre os n primeiros inteiros positivos, uma permutação de grau n pode ser definida como uma função bijetora do conjunto $\{1, 2, \ldots, n\}$ em si mesmo. Tal função σ corresponde à n-upla $(\sigma 1, \sigma 2, \ldots, \sigma n)$ e é simplesmente uma regra para ordenar $1, 2, \ldots, n$ de alguma maneira bem definida.

Se D é uma função n-linear alternada e A é uma $n \times n$ matriz sobre K, então temos

$$D(A) = \sum_{\sigma} A(1, \sigma 1) \ldots A(n, \sigma n)\, D(\varepsilon_{\sigma 1}, \ldots, \varepsilon_{\sigma n}) \tag{5-11}$$

onde a soma é estendida a todas as permutações σ distintas e de grau n.

A seguir mostraremos que

$$D(\varepsilon_{\sigma 1}, \ldots, \varepsilon_{\sigma n} \qquad D(\varepsilon_1, \ldots, \varepsilon_n) \tag{5-12}$$

onde o sinal $\pm$ depende somente da permutação σ. A razão para isto é a que segue. A seqüência $(\sigma 1, \sigma 2, \ldots, \sigma n)$ pode ser obtida da seqüência $(1, 2, \ldots, n)$ por um número finito de transposições de pares de elementos. Por exemplo, se $\sigma 1 \neq 1$, podemos transpor 1 e $\sigma 1$ obtendo $(\sigma 1, \ldots, 1, \ldots)$. Procedendo desta maneira chegaremos à seqüência $(\sigma 1, \ldots, \sigma n)$ após n ou menos transposições de pares. Como D é alternada, o sinal de seu valor muda cada vez que transpomos duas das linhas ε_i e ε_j. Assim, se passamos de $(1, 2, \ldots, n)$ a $(\sigma 1, \sigma 2, \ldots, \sigma n)$ por meio de m transposições de pares (i, j) teremos

$$D(\varepsilon_{\sigma 1}, \ldots, \varepsilon_{\sigma n}) = (-1)^m D(\varepsilon_1, \ldots, \varepsilon_n).$$

Em particular, se D é uma função determinante

$$D(\varepsilon_{\sigma 1}, \ldots, \varepsilon_{\sigma n}) = (-1)^m \tag{5-13}$$

onde m depende somente de σ e não de D. Assim, todas as funções determinantes associam o mesmo valor à matriz com linhas $\varepsilon_{\sigma 1}, \ldots, \varepsilon_{\sigma n}$, e este valor é 1 ou -1.

Um fato básico sobre permutações é o seguinte: se σ é uma permutação de grau n, pode-se passar da seqüência $(1, 2, \ldots, n)$

à seqüência ($\sigma 1, \sigma 2, \ldots, \sigma n$) por meio de uma série de transposições de pares e isto pode ser feito de diversas maneiras; contudo, qualquer que seja a maneira pela qual isto é feito, o número de transposições usadas é sempre par ou sempre ímpar. A permutação é então denominada par ou ímpar, respectivamente. Define-se o sinal de uma permutação por

$$\text{sinal } \sigma = \begin{cases} 1, & \text{se } \sigma \text{ é par} \\ -1, & \text{se } \sigma \text{ é ímpar} \end{cases}$$

com o símbolo "1" indicando aqui o inteiro 1. Mostraremos abaixo que esta propriedade básica das permutações pode ser deduzida do que já sabemos sobre funções determinantes. Suponhamos por ora que isto seja verdade. Então o inteiro m que aparece em (5-13) é sempre par se σ é uma permutação par e é sempre ímpar se σ é uma permutação ímpar. Para qualquer função D n-linear e alternada temos então

$$D(\varepsilon_{\sigma 1}, \ldots, \varepsilon_{\sigma n}) = (\text{sinal } \sigma)\, D(\varepsilon_1, \ldots, \varepsilon_n)$$

e usando (5-11)

$$D(A) = \left[\sum_{\sigma} (\text{sinal } \sigma)\, A(1, \sigma 1) \ldots A(n, \sigma n)\right] D(I). \tag{5-14}$$

É claro que I indica a $n \times n$ matriz unidade.

De (5-14) vemos que existe exatamente uma função determinante sobre as $n \times n$ matrizes sobre K. Se indicarmos esta função por det, ela será dada por

$$\det(A) = \sum_{\sigma} (\text{sinal } \sigma)\, A(1, \sigma 1) \ldots A(n, \sigma n) \tag{5-15}$$

sendo a soma estendida a todas as permutações σ distintas e de grau n. Podemos resumir formalmente como segue.

Teorema 2. *Seja* K *um anel comutativo com elemento unidade e seja* n *um inteiro positivo. Existe exatamente uma função determinante sobre o conjunto das* $n \times n$ *matrizes sobre* K, *que é a função det definida por* (5-15). *Se* D *é uma função* n-*linear alternada arbitrária sobre* $K^{n \times n}$, *então, para toda* $n \times n$ *matriz* A

$$D(A) = (\det A)\, D(I).$$

Este é o teorema que procurávamos, mas deixamos uma lacuna na demonstração. Essa lacuna é a demonstração de que, para uma dada permutação σ, quando passamos de $(1, 2, \dots, n)$ para $(\sigma 1, \sigma 2, \dots, \sigma n)$ transpondo pares, o número de transposições é sempre par ou sempre ímpar. Este fato básico de combinatória pode ser demonstrado sem nenhuma referência a determinantes; contudo, gostaríamos de salientar como ele decorre da *existência* de uma função determinante sobre $n \times n$ matrizes.

Tomemos K como sendo o anel dos inteiros. Seja D uma função determinante sobre as $n \times n$ matrizes sobre K. Seja σ uma permutação de grau n e suponhamos que passemos de $(1, 2, \dots, n)$ a $(\sigma 1, \sigma 2, \dots, \sigma n)$ por meio de m transposições de pares (i, j), $i \neq j$. Como mostramos em (5-13)

$$(-1)^m = D(\varepsilon_{\sigma_1}, \dots, \varepsilon_{\sigma n}),$$

isto é, o número $(-1)^m$ tem que ser o valor de D sobre a matriz de linhas $\varepsilon_{\sigma_1}, \dots, \varepsilon_{\sigma n}$. Se

$$D(\varepsilon_{\sigma_1}, \dots, \varepsilon_{\sigma n}) = 1,$$

então m tem que ser par. Se

$$D(\varepsilon_{\sigma_1}, \dots, \varepsilon_{\sigma n}) = -1,$$

então m tem que ser ímpar.

Como temos uma fórmula explícita para o determinante de uma $n \times n$ matriz e esta fórmula envolve as permutações de grau n, vamos concluir esta seção fazendo mais algumas observações sobre permutações. Primeiro, notemos que existem exatamente $n! = 1 \cdot 2 \dots n$ permutações de grau n, pois se σ é uma tal permutação, existem n escolhas possíveis para $\sigma 1$; uma vez feita, existem $(n-1)$ possibilidades para $\sigma 2'$, depois $(n-2)$ possibilidades para $\sigma 3$ e assim por diante. Logo, existem

$$n(n-1)(n-2) \dots 2 \cdot 1 = n!$$

permutação σ. A fórmula (5-15) para $\det(A)$ fornece desta maneira $\det(A)$ como uma soma de $n!$ termos, um para cada permutação de grau n. Um termo genérico é um produto

$$A(1, \sigma 1) \dots A(n, \sigma n)$$

de n elementos de A, um elemento de cada linha e um de cada coluna, e é acompanhado de um sinal "+" ou "−" conforme seja uma permutação par ou ímpar.

Quando as permutações são consideradas como funções bijetoras do conjunto $\{1, 2, \ldots, n\}$ em si mesmo, é possível definir um produto de permutações. O produto de por τ será simplesmente a função composta $\sigma\tau$ definida por

$$(\sigma\tau)(i) = \sigma(\tau(i)).$$

Se ε indica a permutação idêntica, $\varepsilon(i) = i$, então cada possui uma inversa σ^{-1} tal que

$$\sigma\sigma^{-1} = \sigma^{-1}\sigma = \varepsilon.$$

Pode-se resumir estas observações dizendo que, em relação à operação de composição, o conjunto das permutações de grau n é um grupo. Este grupo é usualmente denominado o **grupo simétrico de grau** n.

Do ponto de vista de produtos de permutações, a propriedade fundamental do sinal de uma permutação é que

(5-16) $$\text{sinal } (\sigma\tau) = (\text{sinal } \sigma)\ (\text{sinal } \tau).$$

Em outras palavras, $\sigma\tau$ é uma permutação par se σ e τ são ambas pares ou ambas ímpares enquanto que $\sigma\tau$ é ímpar se uma das suas permutações é ímpar e a outra é par. Pode-se ver isto pela definição do sinal em termos de permutações sucessivas de pares (i, j). Poderá ser também instrutivo se ressaltarmos como a igualdade sinal $(\sigma\tau) = (\text{sinal } \sigma)$ $(\text{sinal } \tau)$ decorre de uma propriedade fundamental dos determinantes.

Seja K o anel dos inteiros e sejam σ e τ permutações de grau n. Sejam $\varepsilon_1, \ldots, \varepsilon_n$ as linhas da $n \times n$ matriz unidade sobre K, seja A a matriz de linhas $\varepsilon_{\tau 1}, \ldots, \varepsilon_{\tau n}$ e seja B a matriz de linhas $\varepsilon_{\sigma 1}, \ldots, \varepsilon_{\sigma n}$. A i-ésima linha de A contém exatamente um elemento não-nulo, a saber, o 1 na coluna τi. A partir disto, é fácil ver que $\varepsilon_{\sigma\tau i}$ é a i-ésima linha da matriz produto AB. Ora,

$$\det(A) = \text{sinal } \tau, \quad \det(B) = \text{sinal } \sigma \quad \text{e} \quad \det(AB) = \text{sinal } (\sigma\tau).$$

Portanto teremos sinal $(\sigma\tau) = (\text{sinal } \sigma)$ $(\text{sinal } \tau)$ desde que demonstremos o seguinte:

Teorema 3. *Seja* K *um anel comutativo com elemento unidade e sejam* A *e* B n × n *matrizes sobre* K. *Então*

$$det(AB) = (det\ A)\,(det\ B).$$

Demonstração. Seja B uma $n \times n$ matriz fixa sobre K e para cada $n \times n$ matriz A definamos $D(A) = \det(AB)$. Se indicarmos as linhas de A por $\alpha_1, \ldots, \alpha_n$, então

$$D(\alpha_1, \ldots, \alpha_n) = \det(\alpha_1 B, \ldots, \alpha_n B).$$

Aqui $\alpha_j B$ indica a $1 \times n$ matriz que é o produto da $1 \times n$ matriz α_j pela $n \times n$ matriz B. Como

$$(c\alpha_i + \alpha_i')\,B = c\alpha_i B + \alpha_i' B$$

e det é n-linear, é fácil ver que D é n-linear. Se $\alpha_i = \alpha_j$, então, $\alpha_i B = \alpha_j B$, e como det é alternada

$$D(\alpha_1, \ldots, \alpha_n) = 0.$$

Logo, D é alternada. Ora, D é uma função n-linear alternada e pelo Teorema 2

$$D(A) = (\det A)\, D(I).$$

Mas $D(I) = \det(IB) = \det B$, portanto

$$\det\ (AB) = D(A) = (\det A)\,(\det B).$$

O fato de que sinal $(\sigma\tau) = $ (sinal σ) (sinal τ) é apenas um dos muitos corolários do Teorema 3. Consideraremos alguns desses corolários na próxima seção.

Exercícios

1. Se K é um anel comutativo com elemento unidade e A é a matriz sobre K dada por

$$A = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

mostrar que $\det A = 0$.

2. Demonstrar que o determinante da matriz de Vandermonde

$$\begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}$$

é $(b-a)\,(c-a)\,(c-b)$.

3. Enunciar explicitamente as seis permutações de grau 3, dizer quais são ímpares e quais são pares e usar isto para dar uma fórmula completa (5-15) para o determinante de uma 3×3 matriz.

4. Sejam σ e τ as permutações de grau 4 definidas por $\sigma 1 = 2$, $\sigma 2 = 3$, $\sigma 3 = 4$, $\sigma 4 = 1$, $\tau 1 = 3$, $\tau 2 = 1$, $\tau 3 = 2$, $\tau 4 = 4$.

(a) σ é par ou ímpar? τ é par ou ímpar?

(b) Determinar $\sigma\tau$ e $\tau\sigma$.

5. Se A é uma $n \times n$ matriz inversível sobre um corpo, mostrar que $\det A \neq 0$.

6. Seja A uma 2×2 matriz sobre um corpo. Demonstrar que $\det(I + A) = 1 + \det A$ se, e somente se, traço $(A) = 0$.

7. Uma $n \times n$ matriz A é denominada **triangular** se $A_{ij} = 0$ sempre que $i > j$ ou se $A_{ij} = 0$ sempre que $i < j$. Demonstrar que o determinante de uma matriz triangular é o produto $A_{11}A_{22}, \ldots, A_{nn}$ de seus elementos diagonais.

8. Seja A uma 3×3 matriz sobre o corpo dos números complexos. Formemos a matriz $xI - A$ cujos elementos são polinômios, sendo o elemento i, j desta matriz o polinômio $\delta_{ij}x - A_{ij}$. Se $f = \det(xI - A)$, mostrar que f é um polinômio unitário de grau 3. Escrevendo

$$f = (x - c_1)(x - c_2)(x - c_3)$$

com números complexos c_1, c_2, e c_3, demonstrar que

$$c_1 + c_2 + c_3 = \text{traço}(A) \text{ e } c_1c_2c_3 = \det A.$$

9. Seja n um inteiro positivo e F um corpo. Se σ é uma permutação de grau n, demonstrar que a função

$$T(x_1, \ldots, x_n) = (x_{\sigma 1}, \ldots, x_{\sigma n})$$

é um operador linear inversível sobre F^n.

10. Seja F um corpo, n um inteiro positivo e S o conjunto das $n \times n$ matrizes sobre F. Seja V o espaço vetorial de todas as funções de S em F. Seja W o conjunto das funções n lineares alternadas sobre S. Demonstrar que W é um subespaço de V. Qual é a dimensão de W?

11. Seja T um operador linear sobre F^n. Definamos

$$D_T(\alpha_1, \ldots, \alpha_n) = \det(T\alpha_1, \ldots, T\alpha_n)$$

(a) Mostrar que D_T é uma função n-linear alternada.

(b) Se

$$c = \det(T\varepsilon_1, \ldots, T\varepsilon_1)$$

mostrar que para n quaisquer vetores $\alpha_1, \ldots, \alpha_n$ temos

$$\det(T\alpha_1, \ldots, T\alpha_n) = c \det(\alpha_1, \ldots, \alpha_n).$$

(c) Se $\mathscr{B}$ é uma base ordenada arbitrária de F^n e A é a matriz de T em relação à base ordenada $\mathscr{B}$, mostrar que $\det A = c$.

(d) Qual é um nome razoável para o escalar c?

12. Se σ é uma permutação de grau n e A é uma $n \times n$ matriz sobre o corpo F com vetores linhas $\alpha_1, \ldots, \alpha_n$, indiquemos por $\sigma(A)$ a $n \times n$ matriz com vetores-linhas $\alpha_{\sigma 1}, \ldots, \alpha_{\sigma n}$.

(a) Demonstrar que $\sigma(AB) = \sigma(A)B$ e em particular que $\sigma(A) = \sigma(I)A$.

(b) Se T o operador linear do Exercício 9, demonstrar que a matriz T em relação à base ordenada canônica é $\sigma(I)$.

(c) $\sigma^{-1}(I)$ é a matriz inversa de $\sigma(I)$?

(d) É verdade que $\sigma(A)$ é semelhante a A?

13. Demonstrar que a função sinal sobre permutações é única no seguinte sentido: se f é uma função qualquer que associa a cada permutação de grau n um inteiro e se $f(\sigma\tau) = f(\sigma)f(\tau)$, então f é identicamente nula, ou f é identicamente 1 ou f é a função sinal.

5.4 Propriedades Adicionais dos Determinantes

Nesta seção relataremos algumas das propriedades úteis da função determinante sobre $n \times n$ matrizes. Talvez a primeira coisa que devamos destacar seja a que segue. Em nossa discussão de $\det A$, as linhas de A desempenharam um papel privilegiado. Como não existe nenhuma diferença fundamental entre linhas e colunas, pode-se muito bem esperar que $\det A$ seja uma função n-linear alternada das colunas de A. Isto decorre e, para demonstrá-lo, basta mostrar que

$$\det(A^t) = \det(A) \tag{5-17}$$

onde A^t indica a transposta de A.

Se σ é uma permutação de grau n

$$A^t(i, \sigma i) = A(\sigma i, i).$$

Da expressão (5-15) tem-se então

$$\det (A^t) = \sum_\sigma (\text{sinal } \sigma)\, A(\sigma 1, 1) \ldots A(\sigma n, n).$$

Quando $i = \sigma^{-1}j$, $A(\sigma i, i) = A(j, \sigma^{-1}j)$. Assim

$$A(\sigma 1, 1) \ldots A(\sigma n, n) = A(1, \sigma^{-1} 1) \ldots A(n, \sigma^{-1}n).$$

Como $\sigma\sigma^{-1}$ é a permutação idêntica,

$$(\text{sinal } \sigma)\,(\text{sinal } \sigma^{-1}) = 1, \text{ ou seja, } \text{sinal}(\sigma^{-}1) = \text{sinal}(\sigma).$$

Além disso, quando percorre todas as permutações de grau n, σ^{-1} também o faz.

Portanto,

$$\begin{aligned} \det(A^t) &= \sum_\sigma (\text{sinal } \sigma^{-1})\ A(1, \sigma^{-1} 1) \ldots\ A(n, \sigma^{-1}n) \\ &= \det\ A, \end{aligned}$$

o que demonstra (5-17).

Em certas ocasiões é preciso calcular certos determinantes particulares. Quando isto é necessário, é freqüentemente útil tirar vantagem do fato seguinte: *se* B *é obtida de* A *somando-se um múltiplo de uma linha de* A *a outra* (*ou um múltiplo de uma coluna a outra*), *então*

(5-18) $$\det B = \det A.$$

Demonstraremos a afirmação relativa às linhas. Seja B obtida de A somando-se $c\alpha_j$ a α_i, onde $i < j$. Como det é linear como uma função da i-ésima linha

$$\begin{aligned} \det B &= \det A + c \det(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_j, \ldots, \alpha_n) \\ &= \det A. \end{aligned}$$

Outro fato útil é o seguinte: consideremos uma $n \times n$ matriz da forma em blocos

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

onde A é uma $r \times r$ matriz, C é uma $s \times s$ matriz, B é $r \times s$ e 0 indica a $s \times r$ matriz nula. Então

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = (\det A)(\det C). \tag{5-12}$$

Para demonstrar isto, definamos

$$D(A, B, C) = \det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}.$$

Se fixarmos A e B, então D é alternada e s-linear como uma função das linhas de C. Assim, pelo Teorema 2

$$D(A, B, C) = (\det C)\, D(A, B, I)$$

onde I é a $s \times s$ matriz unidade. Subtraindo das linhas de B múltiplos das linhas de I e usando a afirmação (5-18) acima, obtemos

$$D(A, B, I) = D(A, 0, I).$$

Ora, $D(A, 0, I)$ é evidentemente alternada e r-linear como uma função das linhas de A. Assim

$$D(A, 0, I) = (\det A)\, D(I, 0, I).$$

Mas, $D(I, 0, I) = 1$, logo,

$$\begin{aligned} D(A, B, C) &= (\det C)\, D(A, B, I) \\ &= (\det C)\, D(A, 0, I) \\ &= (\det C)(\det A). \end{aligned}$$

Por um raciocínio do mesmo tipo, ou tomando transpostas

$$\det \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} (\det A)(\det C). \tag{5-20}$$

Exemplo 6. Suponhamos que K seja o corpo dos números racionais e que desejamos calcular o determinante de 4×4 matriz

$$A = \begin{bmatrix} 1 & -1 & 2 & 3 \\ 2 & 2 & 0 & 2 \\ 4 & 1 & -1 & -1 \\ 1 & 2 & 3 & 0 \end{bmatrix}.$$

Subtraindo das linhas 2, 3 e 4 múltiplos convenientes da linha 1, obtemos a matriz

$$\begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 5 & -9 & -13 \\ 0 & 3 & 1 & -3 \end{bmatrix}$$

que, como sabemos por (5-18), terá o mesmo determinante que A. Subtraindo da linha 3, $\frac{5}{4}$ da linha 2 e depois subtraindo da linha 4, $\frac{3}{4}$ da linha 2, obtemos

$$B = \begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 0 & -4 & -8 \\ 0 & 0 & 4 & 0 \end{bmatrix}$$

e novamente $\det B = \det A$. A forma em blocos de B nos diz que

$$\det A = \det B = \begin{vmatrix} 1 & -1 \\ 0 & 4 \end{vmatrix} \begin{vmatrix} -4 & -8 \\ 4 & 0 \end{vmatrix} = 4(32) = 128.$$

Seja agora $n > 1$ e seja A uma $n \times n$ matriz sobre K. No Teorema 1 mostramos como construir uma função determinante sobre as $n \times n$ matrizes, dada uma tal função sobre as $(n-1) \times (n-1)$ matrizes. Agora que demonstramos a unicidade da função determinante, a fórmula (5-4) nos diz o que segue. Se fixamos uma coluna arbitrária de índice j,

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det A(i|j).$$

O escalar $(-1)^{i+j} \det A(i|j)$ é usualmente denominado o i, j **cofator** de A ou o cofator do elemento i, j de A. A fórmula acima para $\det A$ é então denominada o desenvolvimento de $\det A$ pelos cofatores da j-ésima coluna (ou, às vezes, o desenvolvimento pelos menores da j-ésima coluna). Se colocarmos

$$C_{ij} = (-1)^{i+j} \det A(i|j)$$

então a fórmula acima dirá que para cada j

$$\det A = \sum_{i=1}^{n} A_{ij} C_{ij}$$

onde o cofator C_{ij} é $(-1)^{i+j}$ vezes o determinante da $(n-1) \times \times (n-1)$ matriz obtida de A retirando-se a i-ésima linha e a j-ésima coluna de A.

Se $j \neq k$, então

$$\sum_{i=1}^{n} A_{ik} C_{ij} = 0.$$

De fato, substituamos a j-ésima coluna de A por sua k-ésima coluna e chamemos a matriz resultante de B. Então B possui duas colunas iguais, logo $\det B = 0$. Como $B(i|j) = A(i|j)$ temos

$$\begin{aligned} 0 &= \det B \\ &= \sum_{i=1}^{n} (-1)^{i+j} B_{ij} \det B(i|j) \\ &= \sum_{i=1}^{n} (-1)^{i+j} A_{ik} \det A(i|j) \\ &= \sum_{i=1}^{n} A_{ik} C_{ij}. \end{aligned}$$

Estas propriedades dos cofatores podem ser resumidas por

(5-21) $$\sum_{i=1}^{n} A_{ik} C_{ij} = \delta_{jk} \det A.$$

A $n \times n$ matriz adj A, que é a transposta da matriz dos cofatores de A, é denominada a **adjunta clássica** de A. Assim

(5-22) $$(\text{adj } A)_{ij} = C_{ji} = (-1)^{i+j} \det A(j|i).$$

As fórmulas (5-21) podem ser resumidas na equação matricial

(5-23) $$(\text{adj } A)\, A = (\det A)\, I.$$

Desejamos também ver que $A(\text{adj } A) = (\det A) I$. Como $A^t(i|j) = = A(j|i)^t$ temos

$$(-1)^{i+j} \det A^t(i|j) = (-1)^{j+i} \det A(j|i)$$

que diz simplesmente que o i, j cofator de A^t é o j, i cofator de A. Assim,

(5-24) $$\text{adj}(A^t) = (\text{adj } A)^t.$$

Aplicando (5-23) a A^t obtemos

$$(\text{adj } A^t)A^t = (\det A^t)I = (\det A)I$$

e transpondo

$$A(\text{adj } A^t)^t = (\det A)I.$$

Usando (5-23) obtemos o que queremos:

(5-25) $$A(\text{adj } A) = (\det A)I.$$

Da mesma forma que para matrizes sobre um corpo, uma $n \times n$ matriz A sobre K é dita **inversível sobre** K se existe uma $n \times n$ matriz A^{-1} com elementos em K tal que $AA^{-1} = A^{-1}A = I$. Se existe uma tal inversa, ela é única, pois o mesmo argumento usado no Capítulo 1 mostra que quando $BA = AC = I$ temos $B = C$. As fórmulas (5-23) e (5-25) nos dizem o seguinte sobre a inversibilidade de matrizes sobre K: se o elemento $\det A$ possui um inverso multiplicativo em K, então A é inversível a $A^{-1} = (\det A)^{-1} \text{ adj } A$ é a única inversa de A. Reciprocamente, é fácil ver que, se A é inversível sobre K, o elemento $\det A$ é inversível em K, pois se $BA = I$ temos

$$1 = \det I = \det(AB) = (\det A)(\det B).$$

O que demonstramos é o seguinte:

Teorema 4. *Seja* A *uma* n × n *matriz sobre* K. *Então* A *é inversível sobre* K *se, e somente se, det* A *é inversível em* K. *Quando* A *é inversível, a única inversa de* A *é*

$$A^{-1} = (\det A)^{-1} \text{ adj } A.$$

Em particular, uma n × n *matriz sobre um corpo é inversível se, e somente se, seu determinante é diferente de zero.*

Gostaríamos de salientar que este critério relativo a determinantes para a inversibilidade demonstra que uma $n \times n$ matriz com

uma inversa à esquerda ou à direita é inversível. Esta demonstração é completamente independente da demonstração que fizemos no Capítulo 1 para matrizes sobre um corpo. Gostaríamos também de ressaltar o que a inversibilidade significa para matrizes cujos elementos são polinômios. Se K é o anel de polinômios $F[x]$, os únicos elementos de K que são inversíveis são os polinômios constantes não-nulos. De fato, se f e g são polinômios e $fg = 1$, temos $\text{gr}(f) + \text{gr}(g) = 0$ de modo que $\text{gr}(f) = \text{gr}(g) = 0$, isto é, f e g são polinômios constantes. Portanto uma $n \times n$ matriz sobre o anel de polinômios $F[x]$ é inversível sobre $F[x]$ se, e somente se, seu determinante é um polinômio constante não-nulo.

Exemplo 7. Seja $K = R[x]$ o anel de polinômios sobre o corpo dos números reais. Sejam

$$A = \begin{bmatrix} x^2 + x & x + 1 \\ x - 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} x^2 - 1 & x + 2 \\ x^2 - 2x - 3 & x \end{bmatrix}.$$

Então, por meio de cálculos simples, $\det A = x + 1$ e $\det B = -6$. Assim, A não é inversível sobre K, enquanto que B é inversível sobre K. Notemos que

$$\text{adj } A = \begin{bmatrix} 1 & -x - 1 \\ -x + 1 & x^2 + x \end{bmatrix}, \quad \text{adj } B = \begin{bmatrix} x & -x - 2 \\ -x^2 + 2x - 3 & x^2 - 1 \end{bmatrix}$$

e $(\text{adj } A)A = (x + 1)I$, $(\text{adj } B)B = -6I$. É claro que

$$B^{-1} = -\frac{1}{6}\begin{bmatrix} x & -x - 2 \\ -x^2 + 2x - 3 & 1 - x^2 \end{bmatrix}$$

Exemplo 8. Seja **K** o anel dos inteiros e

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Então $\det A = -2$ e

$$\text{adj } A = \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}.$$

Assim, A não é inversível como uma matriz sobre o anel dos inteiros; no entanto, podemos também considerar A como uma matriz

sobre o corpo dos números racionais. Se o fazemos, então A é inversível e

$$A^{-1} = -\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}.$$

Com relação a matrizes inversíveis, gostaríamos de mencionar mais um fato elementar. Matrizes semelhantes têm o mesmo determinante, isto é, se P é inversível sobre K e $B = P^{-1}AP$, então $\det B = \det A$. Isto é evidente, pois

$$\det(P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A.$$

Esta observação simples torna possível definir o determinante de um operador linear sobre um espaço vetorial de dimensão finita. Se T é um operador linear sobre V, definimos o determinante de T como sendo o determinante de qualquer $n \times n$ matriz que represente T em relação a alguma base ordenada de V. Como todas essas matrizes são semelhantes, elas possuem o mesmo determinante e nossa definição faz sentido. Em relação a isto, veja o Exercício 11 da Seção 5-3.

Gostaríamos agora de discutir a **regra de Cramer** para a resolução de sistemas de equações lineares. Suponhamos que A seja uma $n \times n$ matriz sobre o corpo F e que desejamos resolver o sistema de equações lineares $AX = Y$ para uma dada n-upla $(y_1, \ldots, y_n)$. Se $AX = Y$, então

$$(\text{adj } A)AX = (\text{adj } A)Y$$

e, portanto,

$$(\det A)X = (\text{adj } A)Y.$$

Assim

$$\begin{aligned}(\det A)x_j &= \sum_{i=1}^{n} (\text{adj } A)_{ji}y_i \\ &= \sum_{i=1}^{n} (-1)^{i+j}y_i \det A(i|j).\end{aligned}$$

Esta última expressão é o determinante de $n \times n$ matriz obtida substituindo a j-ésima coluna de A por Y. Se $\det A = 0$, tudo isto não nos diz nada; contudo, se $\det A \neq 0$ temos o que é conhecido como a regra de Cramer. Seja A uma $n \times n$ matriz sobre o corpo F tal

que $\det A \neq 0$. Se $y_1, \ldots, y_n$ são escalares arbitrários em F, a única solução $X = A^{-1}Y$ do sistema de equações $AX = Y$ é dada por

$$x_j = \frac{\det B_j}{\det A}, \qquad j = 1, \ldots, n$$

onde B_j é a $n \times n$ matriz obtida de A substituindo a j-ésima coluna de A por Y.

Concluindo esta seção, gostaríamos de fazer alguns comentários que sirvam para colocar os determinantes naquilo que acreditamos ser a perspectiva apropriada. De vez em quando, é necessário calcular alguns determinantes particulares e esta seção foi parcialmente dedicada a técnicas que irão facilitar esse trabalho. No entanto, o papel principal dos determinantes neste livro é teórico. Não se discute a beleza de fatos tais como a regra de Cramer. Mas a regra de Cramer é um instrumento ineficiente para resolver sistemas de equações lineares, principalmente porque envolve cálculos em demasia. Portanto, deve-se concentrar no que a regra de Cramer diz, e não em como calcular por meio dela. De fato, refletindo sobre este capítulo, esperamos que o leitor coloque mais ênfase na compreensão do que a função determinante é e como ela se comporta do que no modo de calcular determinantes de matrizes particulares.

Exercícios

1. Usar a fórmula da adjunta clássica para calcular as inversas das 3×3 matrizes reais seguintes:

$$\begin{bmatrix} -2 & 3 & 2 \\ 6 & 0 & 3 \\ 4 & 1 & -1 \end{bmatrix} \qquad \begin{bmatrix} \cos\theta & 0 & -\operatorname{sen}\theta \\ 0 & 1 & 0 \\ \operatorname{sen}\theta & 0 & \cos\theta \end{bmatrix}$$

2. Usar a regra de Cramer para resolver os seguintes sistemas de equações lineares sobre o corpo dos números racionais:

(a)
$$\begin{aligned} x + y + z &= 11 \\ 2x - 6y - z &= 0 \\ 3x + 4y + 2z &= 0 \end{aligned}$$

(b)
$$\begin{aligned} 3x - 2y &= 7 \\ 3y - 2z &= 6 \\ 3z - 2x &= -1 \end{aligned}$$

3. Uma $n \times n$ matriz A sobre um corpo F é **anti-simétrica** se $A^t = -A$. Se A é uma $n \times n$ matriz anti-simétrica com elementos complexos e se n é ímpar, demonstrar que $\det A = 0$.

4. Uma $n \times n$ matriz A sobre um corpo F é dita **ortogonal** se $AA^t = I$. Se A é ortogonal, mostrar que $\det A = \pm 1$. Dar um exemplo de uma matriz ortogonal para a qual $\det A = -1$.

5. Uma $n \times n$ matriz A sobre o corpo dos números complexos é dita **unitária** se $AA^* = I$ (A^* indica a transposta conjugada de A). Se A é unitária, mostrar que $|\det A| = 1$.

6. Sejam T e U operadores lineares sobre o espaço vetorial V de dimensão finita. Demonstrar que
   (a) $\det(TU) = (\det T)(\det U)$;
   (b) T é inversível se, e somente se, $\det T \neq 0$.

7. Seja A uma $n \times n$ matriz sobre K, um anel comutativo com elemento unidade. Suponhamos que A seja da forma em blocos

$$A = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & A_k \end{bmatrix}$$

onde A_j é uma $r_j \times r_j$ matriz. Demonstrar que

$$\det A = (\det A_1)(\det A_2) \dots (\det A_k).$$

8. Seja V o espaço vetorial das $n \times n$ matrizes sobre o corpo F. Seja B um elemento fixo de V e seja T_B o operador linear sobre V definido por $T_B(A) = AB - BA$. Mostrar que $\det T_B = 0$.

9. Seja A uma $n \times n$ matriz sobre um corpo. $A \neq 0$. Se r é um inteiro positivo arbitrário entre 1 e n, uma $r \times r$ **submatriz de** A é qualquer $r \times r$ matriz obtida retirando $(n-r)$ linhas e $(n-r)$ colunas de A. O **posto-determinante** de A é o maior inteiro positivo r tal que alguma $r \times r$ submatriz de A possua um determinante não-nulo. Demonstrar que o posto-determinante de A é igual ao posto-linha de A (= posto-coluna de A).

10. Seja A uma $n \times n$ matriz sobre o corpo F. Demonstrar que existem no máximo n escalares c distintos em F tais que $\det(cI - A) = 0$.

11. Sejam A e B $n \times n$ matrizes sobre o corpo F. Mostrar que se A é inversível existem no máximo n escalares c em F para os quais a matriz $cA + B$ não é inversível.

12. Seja V o espaço vetorial das $n \times n$ matrizes sobre F, B uma $n \times n$ matriz fixa sobre F e sejam L_B e R_B os operadores lineares sobre V definidos por $L_B(A) = BA$ e $R_B(A) = AB$. Mostrar que
   (a) $\det L_B = (\det B)^n$;
   (a) $\det R_B = (\det B)^n$.

13. Seja V o espaço vetorial das $n \times n$ matrizes sobre o corpo dos números complexos e seja B uma $n \times n$ matriz sobre C, fixa. Definamos um operador linear M_B sobre V por $M_B(A) = BAB^*$, onde $B^* = \bar{B}^t$. Mostrar que

$$\det M_B = |\det B|^{2n}.$$

Seja agora H o conjunto de todas as matrizes hermitianas em V. (A é hermitiana se $A = A^*$.) Então H é um espaço vetorial sobre o corpo dos números *reais*. Mostrar que a função T_B definida por $T_B(A) = BAB^*$ é um operador linear sobre o espaço vetorial real H e depois mostrar que $T_B = |\det B|^{2n}$. (*Sugestão*: Ao calcular det T_B, mostrar que V possui uma base constituída de matrizes hermitianas e então mostrar que $\det T_B = \det M_B$.)

14. Sejam A, B, C, D $n \times n$ matrizes sobre o corpo F tais que elas comutam. Mostrar que o determinante da $2n \times 2n$ matriz

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

é $\det(AD - BC)$.

5.5 Módulos

Se K é um anel comutativo com elemento unidade, um módulo sobre K é um sistema algébrico que se comporta como um espaço vetorial, K fazendo o papel do corpo dos escalares. Para sermos precisos, diremos que V é um **módulo sobre** K (ou um K-**módulo**) se

1. existe uma adição $(\alpha, \beta) \to \alpha + \beta$ em V, em relação à qual V é um grupo comutativo;

2. existe uma multiplicação $(c, \alpha) \to c\alpha$ de elementos α em V e c em K, tal que

$$(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$$
$$c(\alpha_1 + \alpha_2) = c\alpha_1 + c\alpha_2$$
$$(c_1c_2)\alpha = c_1(c_2\alpha)$$
$$1\alpha = \alpha.$$

Para nós, os K-módulos de maior importância serão os módulos K^n, das n-uplas. Os módulos $K^{m \times n}$, das matrizes, também são importantes. Se V é um módulo qualquer, falaremos de combinações lineares, dependência linear, independência linear, exatamente como num espaço vetorial. Devemos ter o cuidado de não aplicar a V qualquer resultado de espaços vetoriais que dependa da divisão por escalares não nulos, a única operação de corpo que

poderia estar ausente no anel K. Por exemplo, se $\alpha_1, \ldots, \alpha_n$ são linearmente dependentes, não poderemos concluir que algum α_i seja uma combinação linear dos demais. Isto torna mais difícil determinar bases em módulos.

Uma base do módulo V é um conjunto linearmente independente que gera o módulo. Esta definição é igual à que foi dada para espaços vetoriais; e a propriedade importante de uma base $\mathcal{B}$, é que todo elemento de V pode ser escrito, de modo único, como uma combinação linear de (algum número finito de) elementos de $\mathcal{B}$. Se admitirmos na matemática o Axioma da Escolha (veja o Apêndice), poderemos mostrar que todo espaço vetorial possui uma base. O leitor está bem ciente que existe uma base em qualquer espaço vetorial que seja gerado por um número finito de vetores. Isto não acontece em módulos. Por isso precisamos de denominações especiais para módulos que possuam bases e para módulos gerados por um número finito de elementos.

Definição. *O* K*-módulo* V *é denominado* **módulo livre** *se ele possuir uma base. Se* V *possuir uma base contendo* n *elementos,* V *é denominado* K-**módulo livre com** n **geradores.**

Definição. *O módulo* V *é* **de tipo finito** *se ele contiver um subconjunto finito que gere* V. *O* **posto** *de um módulo de tipo finito é o menor inteiro* k *tal que* k *elementos gerem* V.

Repetimos que um módulo pode ser de tipo finito sem que tenha uma base finita. Se V é um K-módulo livre com n geradores, então V é isomorfo ao módulo K^n. Se $\{\beta_1, \ldots, \beta_n\}$ é uma base de V, existe um isomorfismo que leva o vetor $c_1\beta_1 + \ldots + c_n\beta_n$ na n-upla $(c_1, \ldots, c_n)$ de K^n. Não é aparente, de imediato, que o mesmo módulo V não possa também ser um módulo livre com k geradores, $k \neq n$. Em outras palavras, não é evidente que duas bases quaisquer de V devam conter o mesmo número de elementos. A demonstração deste fato é uma aplicação interessante de determinantes.

Teorema 5. *Seja* K *um anel comutativo com elemento unidade. Se* V *é um* K*-módulo livre com* n *geradores, então o posto de* V *é* n.

Demonstração. Devemos provar que V não pode ser gerado por menos de n dos seus elementos. Sendo V isomorfo a K^n, devemos mostrar que se $m < n$, o módulo K^n não é gerado por n-uplas $\alpha_1, \ldots, \alpha_m$. Seja A a matriz com linhas $\alpha_1, \ldots, \alpha_m$. Suponhamos que

cada um dos vetores da base canônica $\varepsilon_1, \ldots, \varepsilon_n$ seja uma combinação linear de $\alpha_1, \ldots, \alpha_m$. Então existe uma matriz P, em $K^{n \times m}$, tal que

$$PA = I,$$

onde I é a $n \times n$ matriz unidade. Seja $\tilde{A}$ a $n \times n$ matriz que se obtém acrescentando n-m linhas de O's à matriz A e seja $\tilde{P}$ uma $n \times n$ matriz arbitrária, cujas primeiras n colunas sejam as colunas de P. Então

$$\tilde{P}\tilde{A} = I.$$

Portanto, det $\tilde{A} \neq 0$. Sendo, porém, $m < n$, pelo menos uma linha de $\tilde{A}$ é constituída somente de O. Esta contradição mostra que K^n não é gerado por $\alpha_1, \ldots, \alpha_m$.

É interessante notar que o Teorema 5 estabelece a unicidade da dimensão de um espaço vetorial (de dimensão finita). A demonstração, baseada na existência da função determinante, é bem diferente da demonstração dada no Capítulo 2. Pelo Teorema 5 sabemos que "módulo livre de posto n" é o mesmo que "módulo livre com n geradores".

Se V é um módulo sobre K, o **módulo dual** V^* consiste de todas as funções lineares f de V em K. Se V é um módulo livre de posto n, então V^* é também um módulo livre de posto n. A demonstração é igual à que foi feita para espaços vetoriais. Se $\{\beta_1, \ldots, \beta_n\}$ é uma base ordenada de V, existe uma **base dual** associada, $\{f_1, \ldots, f_n\}$ para o módulo V^*. A função f_i associa a cada α em V sua i-ésima coordenada em relação a $\{\beta_1, \ldots, \beta_n\}$:

$$\alpha = f_1(\alpha)\beta_1 + \ldots + f_n(\alpha)\beta_n.$$

Se f é uma função linear sobre V,

$$f = f(\beta_1)f_1 + \ldots + f(\beta_n)f_n.$$

5.6 Funções Multilineares

O objetivo desta seção é colocar nossa discussão sobre determinantes no que acreditamos ser a perspectiva apropriada. Trataremos de formas multilineares alternadas sobre módulos. Estas

formas são a generalização natural dos determinantes, como foram apresentados. O leitor que não leu (ou não tiver desejo de ler) a rápida descrição de módulos da Seção 5.5, ainda pode estudar esta seção com proveito, substituindo, consistentemente, a expressão "módulo livre sobre K de posto n" por "espaço vetorial sobre F de dimensão n".

Seja K um anel comutativo com elemento unidade e seja V um módulo sobre K. Se r é um inteiro positivo, a função L de $V^r = V \times V \times \ldots \times V$ em K é denominada **multilinear** se $L(\alpha_1, \ldots, \alpha_r)$ é linear, como uma função de cada α_i quando os demais α_j são mantidos fixos, isto é, se para cada i

$$L(\alpha_1, \ldots, c\alpha_i + \beta_i, \ldots, \alpha_r) = cL(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_r + \\ + L(\alpha_1, \ldots, \beta_i, \ldots, \alpha_r).$$

Uma função multilinear sobre V^r será também denominada **forma r-linear** sobre V ou **forma multilinear de grau** r sobre V. Tais funções são às vezes chamadas **r-tensores** sobre V. A coleção de todas as funções multilineares sobre V^r será indicada por $M^r(V)$. Se L e M estão em $M^r(V)$, a soma $L + M$:

$$(L + M)(\alpha_1, \ldots, \alpha_r) = L(\alpha_1, \ldots, \alpha_r) + M(\alpha_1, \ldots, \alpha_r)$$

é também multilinear; e, sendo c um elemento de K, o produto cL:

$$(cL)(\alpha_1, \ldots, \alpha_r) = cL(\alpha_1, \ldots, \alpha_r)$$

é multilinear. Portanto, $M^r(V)$ é um K-módulo — um submódulo do módulo de todas as funções V^r em K.

Se $r = 1$, temos $M^1(V) = V^*$, o módulo dual das funções lineares sobre V. As funções lineares também podem ser usadas para construir exemplos de formas multilineares de grau mais elevado. Se $f_1, \ldots, f_r$ são funções lineares sobre V, definamos

$$L(\alpha_1, \ldots, \alpha_r) = f_1(\alpha_1) f_2(\alpha_2) \ldots f_r(\alpha_r).$$

Evidentemente L é uma forma r-linear sobre V.

Exemplo 9. Se V é um módulo, uma forma 2-linear sobre V é denominada **forma bilinear** sobre V. Seja A uma $n \times n$ matriz com elementos em K. Então,

$$L(X, Y) = Y^t AX$$

define uma forma bilinear sobre o módulo $K^{n \times 1}$. Analogamente

$$M(\alpha, \beta) = \alpha A \beta^t$$

define uma forma bilinear M sobre K^n.

Exemplo 10. A função determinante associa a cada $n \times n$ matriz A, um elemento det A, em K. Se considerarmos det A como uma função das linhas de A:

$$\det A = D(\alpha_1, \ldots, \alpha_n)$$

então D é uma forma n-linear sobre K^n.

Exemplo 11. É fácil obter uma expressão algébrica para a forma r-linear, geral, sobre o módulo K^n. Se $\alpha_1, \ldots, \alpha_r$ são vetores em V e A é a $r \times n$ matriz com linhas $\alpha_1, \ldots, \alpha_r$, então para qualquer função L em $M^r(K^n)$,

$$\begin{aligned} L(\alpha_1, \ldots, \alpha_r) &= L\left(\sum_{j=1}^{n} A_{1j}\varepsilon_j, \alpha_2, \ldots, \alpha_r\right) \\ &= \sum_{j=1}^{n} A_{1j} L(\varepsilon_j, \alpha_2, \ldots, \alpha_r) \\ &= \sum_{j=1}^{n} A_{1j} L\left(\varepsilon_j, \sum_{j=1}^{n} A_{2k}\varepsilon_k, \ldots, \alpha_r\right) \\ &= \sum_{j=1}^{n} \sum_{k=1}^{n} A_{1j} A_{2k} L(\varepsilon_j, \varepsilon_k, \alpha_3, \ldots, \alpha_r) \\ &= \sum_{j,k=1}^{n} A_{1j} A_{2k} L(\varepsilon_j, \varepsilon_k, \alpha_3, \ldots, \alpha_r). \end{aligned}$$

Se, por sua vez, substituirmos $\alpha_1, \ldots, \alpha_r$ pelas suas expressões como combinações lineares dos vetores da base canônica, e escrevermos $A(i, j)$ no lugar de A_{ij}, obteremos o seguinte:

$$(5\text{-}26) \qquad L(\alpha_1, \ldots, \alpha_r) = \sum_{j_1, \ldots, j_r = 1}^{n} A(1, j_1) \ldots A(r, j_r)\, L(\varepsilon_{j_1}, \ldots \varepsilon_{j_r}).$$

Em (5-26) existe um termo para cada r-upla $J = (j_1, \ldots, j_r)$ de inteiros positivos entre 1 e n. Existem n^r r-uplas deste tipo. Logo, L é completamente determinada por (5-26) e os valores particulares

$$c_J = L(\varepsilon_{j_1}, \ldots, \varepsilon_{j_r})$$

atribuídos aos n^r elementos $(\varepsilon_{j_1}, \ldots, \varepsilon_{j_r})$. É também fácil ver que escolhendo para cada r-upla J um elemento c_J de K

$$L(\alpha_1, \ldots, \alpha_r) = \sum_J A(1, j_1) \ldots A(r, j_r)\, c_J \tag{5-27}$$

define uma forma r-linear sobre K^n.

Suponhamos que L seja uma função multilinear sobre V^r e M uma função multilinear sobre V^s. Definamos uma função $L \otimes M$ sobre V^{r+s} por

$$(L \otimes M)(\alpha_1, \ldots, \alpha_{r+s}) = L(\alpha_1, \ldots, \alpha_r)\, M(\alpha_{r+1}, \ldots, \alpha_{r+s}). \tag{5-28}$$

Se considerarmos V^{r+s} como sendo $V^r \times V^s$, então para α em V^r e β em V^s

$$(L \otimes M)(\alpha, \beta) = L(\alpha)\, M(\beta).$$

É claro que $L \otimes M$ é multilinear sobre V^{r+s}. A função $L \otimes M$ é denominada **produto tensorial** de L e M. O produto tensorial não é comutativo. De fato, $M \otimes L \neq L \otimes M$ a não ser quando $L = 0$ ou $M = 0$; no entanto, o produto tensorial está intimamente relacionado com as operações de módulo em M^r e M^s.

Lema. *Sejam* L *e* L_1 *formas* r-*lineares sobre* V, M *e* M_1 *formas* s-*lineares sobre* V *e* c *um elemento de* K.

(a) $(cL + L_1) \otimes M = c(L \otimes M) + L_1 \otimes M$;
(b) $L \otimes (cM + M_1) = c(L \otimes M) + L \otimes M_1$.

Demonstração. Exercício.

O produto tensorial é associativo, isto é, se L, M e N são formas (respectivamente) r-, s- e t-lineares sobre V, então

$$(L \otimes M) \otimes N = L \otimes (M \otimes N).$$

Isto decorre imediatamente do fato de que a multiplicação em K é associativa. Portanto, se $L_1, L_2, \ldots, L_k$ são funções multilineares sobre $V^{r_1}, \ldots, V^{r_k}$, o produto tensorial

$$L = L_1 \otimes \ldots \otimes L_k$$

é definido, sem ambigüidade, como uma função multilinear sobre V^r, sendo $r = r_1 + \ldots + r_k$. Um caso particular disto foi mencionado anteriormente. Se $f_1, \ldots, f_r$ são funções lineares sobre V, o produto tensorial

$$L = f_1 \otimes \ldots \otimes f_r$$

é dado por

$$L(\alpha_1, \ldots, \alpha_r) = f_1(\alpha_1) \ldots f_r(\alpha_r).$$

Teorema 6. *Seja* K *um anel comutativo com elemento unidade. Se* V *é um* K*-módulo livre de posto* n, *então* $\mathrm{M}^r(\mathrm{V})$ *é um* K*-módulo livre de posto* n^r; *de fato, se* $\{\mathrm{f}_1, \ldots, \mathrm{f}_n\}$ *é uma base dual* V*, *os* n^r produtos tensoriais

$$\mathrm{f}_{j_1} \otimes \ldots \otimes \mathrm{f}_{j_r}, \qquad 1 \leq \mathrm{j}_1 \leq \mathrm{n}, \ldots, 1 \leq \mathrm{j}_r \leq \mathrm{n}$$

formam uma base de $\mathrm{M}^r(\mathrm{V})$.

Demonstração. Seja $\{f_1, \ldots, f_n\}$ uma base ordenada de V^* e que seja dual da base $\{\beta_1, \ldots, \beta_n\}$ de V. Para cada vetor α em V, temos

$$\alpha = f_1(\alpha)\beta_1 + \ldots + f_n(\alpha)\beta_n.$$

Efetuamos agora o cálculo desenvolvido no Exemplo 11. Se L é uma forma r-linear sobre V e $\alpha_1, \ldots, \alpha_r$ são elementos de V, então de (5-26)

$$L(\alpha_1, \ldots, \alpha_r) = \sum_{j_1, \ldots, j_r} f_{j_1}(\alpha_1) \ldots f_{j_r}(\alpha_r) L(\beta_{j_1}, \ldots, \beta_{j_r}).$$

Em outras palavras,

(5-29) $$L = \sum_{j_1, \ldots, j_r} L(\beta_{j_1}, \ldots, \beta_{j_r}) f_{j_1} \otimes \ldots \otimes f_{j_r}.$$

Isto mostra que os n^r produtos tensoriais

(5-30) $$E_J = f_{j_1} \otimes \ldots \otimes f_{j_r}$$

dados pelas r-uplas $J = (j_1, \ldots, j_r)$ geram o módulo $M^r(V)$. Vemos que as várias r-formas E_J são independentes, da seguinte maneira: suponhamos que para cada J tenha-se um elemento c_J em K e formemos a função multilinear

$$L = \sum_J c_J E_J. \tag{5-31}$$

Observemos que se $I = (i_1, \ldots, i_r)$, então

$$E_J(\beta_{i_1}, \ldots, \beta_{i_r}) = \begin{cases} 0, & I \neq J \\ 1, & I = J. \end{cases}$$

Portanto, utilizando (5-31), temos

$$c_I = L(\beta_{i_1}, \ldots, \beta_{i_r}). \tag{5-32}$$

Em particular, se $L = 0$ então $c_I = 0$ para cada r-upla I.

Definição. *Seja* L *uma forma* r*-linear sobre um* K*-módulo* V. *Dizemos que* L *é alternada se* $L(\alpha_1, \ldots, \alpha_r) = 0$ *sempre que* $\alpha_i = \alpha_j$ *com* $i \neq j$.

Se L é uma função multilinear alternada sobre V^r,

$$L(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_j, \ldots, \alpha_r) = -L(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_i, \ldots, \alpha_r).$$

Em outras palavras, transpondo dois dos vetores (de índices diferentes) na r-upla $(\alpha_1, \ldots, \alpha_r)$ o valor correspondente de L muda de sinal. Como toda permutação é um produto de transposições, vemos que $L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma r}) = (\text{sinal } \sigma)\ L(\alpha_1, \ldots, \alpha_r)$.

Indicaremos por $\Lambda^r(V)$ a coleção de todas as formas r-lineares alternadas sobre V. É claro que $\Lambda^r(V)$ é um submódulo de $M^r(V)$.

Exemplo 12. Mostramos anteriormente, neste capítulo, que existe exatamente uma forma n-linear alternada D, sobre o módulo K^n, tal que $D(\varepsilon_1, \ldots, \varepsilon_n) = 1$. Mostramos também, no Teorema 2, que se L é uma forma arbitrária em $\Lambda^n(K^n)$,

$$L = L(\varepsilon_1, \ldots, \varepsilon_n) D.$$

Em outras palavras, $\Lambda^n(K^n)$ é um K-módulo livre de posto 1. Em (5-15) desenvolvemos uma fórmula explícita para D. Com a notação que estamos usando agora, esta fórmula pode ser escrita

(5-33) $$D = \sum_\sigma (\text{sinal } \sigma) f_{\sigma 1} \otimes \dots \otimes f_{\sigma n}.$$

onde $f_1, \dots, f_n$ são as funções coordenadas canônicas sobre K^n e a somatória se estende sobre as $n!$ permutações distintas σ do conjunto $\{1, \dots, n\}$. Se indicarmos o determinante de uma matriz A por

$$\det A = \sum_\sigma (\text{sinal } \sigma) A(\sigma 1, 1) \dots A(\sigma n, n)$$

obteremos uma expressão diferente para D:

(5-34) $$\begin{aligned} D(\alpha_1, \dots, \alpha_n) &= \sum_\sigma (\text{sinal } \sigma) f_1(\alpha_{\sigma 1}) \dots f_n(\alpha_{\sigma n}) \\ &= \sum_\sigma (\text{sinal } \sigma) L(\alpha_{\sigma 1}, \dots, \alpha_{\sigma n}) \end{aligned}$$

onde

$$L = f_1 \otimes \dots \otimes f_n.$$

Existe um método geral para associar uma forma multilinear com uma forma alternada. Se L é uma forma r-linear sobre um módulo V e se σ é uma permutação de $\{1, \dots, r\}$ obtemos uma outra função L_σ, r-linear, definindo

$$L_\sigma(\alpha_1, \dots, \alpha_r) = L(\alpha_{\sigma 1}, \dots, \alpha_{\sigma r}).$$

Se L for alternada então $L_\sigma = (\text{sinal } \sigma) L$. Agora, para cada L em $M^r(V)$ definamos uma função $\pi_r L$ em $M^r(V)$ por:

(5-35) $$\pi_r L = \sum_\sigma (\text{sinal } \sigma) L_\sigma$$

isto é,

(5-36) $$(\pi_r L)(\alpha_1, \dots, \alpha_r) = \sum_\sigma (\text{sinal } \sigma) L(\alpha_{\sigma 1}, \dots, \alpha_{\sigma r}).$$

Lema. π_r *é uma transformação linear de* $M^r(V)$ *em* $\Lambda^r(V)$. *Se* L *está em* $\Lambda^r(V)$ *então* $\pi_r L = r!\, L$.

Demonstração. Seja τ uma permutação arbitrária de $\{1, \dots, r\}$. Então

$$\begin{aligned} (\pi_r L)(\alpha_{\tau 1}, \dots, \alpha_{\tau r}) &= \sum_\sigma (\text{sinal } \sigma) L(\alpha_{\tau\sigma 1}, \dots, \alpha_{\tau\sigma r}) \\ &= (\text{sinal } \tau) \sum_\sigma (\text{sinal } \tau\sigma) L(\alpha_{\tau\sigma 1}, \dots, \alpha_{\tau\sigma r}). \end{aligned}$$

Quando σ percorre (uma vez) todas as permutações de $\{1, \ldots, r\}$ o mesmo acontece com $\tau\sigma$. Portanto,

$$(\pi_r L)(\alpha_{r1}, \ldots, \alpha_{rr}) = (\text{sinal } \tau)(\pi_r L)(\alpha_1, \ldots, \alpha_r).$$

Logo, $\pi_r L$ é uma forma alternada.

Se L está em $\Lambda^r(V)$, então $L(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma r}) = (\textit{sinal } \sigma)\, L(\alpha_1, \ldots, \alpha_r)$ para cada σ; logo, $\pi_r L = r!\, L$.

Mostramos, em (5-33), que a função determinante D em $\Lambda^n(K^n)$ é

$$D = \pi_n(f_1 \otimes \ldots \otimes f_n)$$

onde $f_1, \ldots, f_n$ são as funções coordenadas canônicas sobre K^n. Existe uma observação importante que deve ser feita em relação a este último lema. Se K é um corpo tal que $r!$ seja inversível em K, então π leva $M^r(V)$ sobre $\Lambda^r(V)$. De fato, neste caso, é mais natural, de um certo ponto de vista, usar a aplicação $\pi_1 = (1/r!)\pi$ no lugar de π, pois π_1 é a projeção de $M^r(V)$ sobre $\Lambda^r(V)$, isto é, π_1 é uma aplicação linear de $M^r(V)$ sobre $\Lambda^r(V)$ tal que $\pi_1(L) = L$ se, e somente se, L estiver em $\Lambda^r(V)$.

Teorema 7. *Seja* K *um anel comutativo com elemento unidade e seja* V *um* K-*módulo livre de posto* n. *Se* $r > n$, $\Lambda^r(V) = \{0\}$. *Se* $1 \leq r \leq n$, *então* $\Lambda^r(V)$ *é um* K-*módulo livre de posto* $\binom{n}{r}$.

Demonstração. Seja $\{\beta_1, \ldots, \beta_n\}$ uma base ordenada de V com base dual $\{f_1, \ldots, f_n\}$. Se L está em $M^r(V)$, temos

$$L = \sum_J L(\beta_{j1}, \ldots, \beta_{jr})\, f_{j1} \otimes \ldots \otimes f_{jr} \tag{5-37}$$

onde a somatória se estende sobre todas as r-uplas $J = (j_1, \ldots, j_r)$ de inteiros entre 1 e n. Se L é alternada,

$$L(\beta_{j1}, \ldots, \beta_{jr}) = 0$$

sempre que dois dos índices j_i forem iguais. Se $r > n$, em cada r-upla J, algum inteiro é necessariamente repetido. Portanto, $\Lambda^r(V) = \{0\}$ se $r > n$.

Suponhamos agora $1 \leq r \leq n$. Se L está em $\Lambda^r(V)$, a soma em (5-37) precisa ser efetuada apenas para as r-uplas J para as quais

$j_1, \ldots, j_r$ são distintos, uma vez que todos os outros termos são nulos. Cada r-upla de inteiros distintos entre 1 e n é uma permutação de uma r-upla $J = (j_1, \ldots, j_r)$ tal que $j_1 < \ldots < j_r$. Este tipo especial de r-upla é denominado r-upla crescente de elementos de $\{1, \ldots, n\}$. Existem

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

tais r-uplas.

Suponhamos fixado uma r-upla crescente J. Seja L_J a soma de todos os termos em (5-37) correspondentes a permutações da r-upla J. Se σ é uma permutação de $\{1, \ldots, r\}$, então

$$L(\beta_{j\sigma 1}, \ldots, \beta_{j\sigma r}) = (\text{sinal } \sigma)\, L(\beta_{j1}, \ldots, \beta_{jr}).$$

Assim,

$$L_J = L(\beta_{j1}, \ldots, \beta_{jr}) D_J \tag{5-38}$$

onde

$$\begin{aligned} D_J &= \sum_{\sigma} (\text{sinal } \sigma) f_{j\sigma 1} \otimes \ldots \otimes f_{j\sigma r} \\ &= \pi_r(f_{j1} \otimes \ldots \otimes f_{jr}). \end{aligned} \tag{5-39}$$

De (5-39) vemos que cada D_J é alternada e que

$$L = \sum_J L(\beta_{j1}, \ldots, \beta_{jr}) D_J \tag{5-40}$$

para toda L em $\Lambda^r(V)$, onde o índice da somatória percorre o conjunto de todas as r-uplas crescentes. Afirmamos que as $\binom{n}{r}$ formas D_j constituem uma base de $\Lambda^r(V)$. Vimos que elas geram $\Lambda^r(V)$. É fácil ver que elas são independentes, da seguinte maneira: se $I = (i_1, \ldots, i_r)$ e $J = (j_1, \ldots, j_r)$ são r-uplas crescentes, então

$$D_J(\beta_{i_1}, \ldots, \beta_{i_r}) = \begin{cases} 1, & I = J \\ 0, & I \neq J \end{cases}. \tag{5-41}$$

Suponhamos que c_J seja um escalar associado a cada r-upla crescente, e definamos

$$L = \sum_J c_J D_J.$$

De (5-40) e (5-41) obtemos

$$c_I = L(\beta_{i_1}, \ldots, \beta_{i_r}).$$

Em particular, se $L = 0$, $c_I = 0$ para cada r-upla crescente I.

Corolário. *Se* V *é um* K-*módulo livre de posto* n, *então* $\Lambda^r(V)$ *é um* K-*módulo livre de posto* 1. *Se* T *é um operador linear sobre* V, *então existe um único elemento* c *em* K, tal que

$$L(T\alpha_1, \ldots, T\alpha_n) = cL(\alpha_1, \ldots, \alpha_n)$$

para toda forma L, n-*linear alternada, sobre* V.

Demonstração. Se L está em $\Lambda^n(V)$, evidentemente

$$L_T(\alpha_1, \ldots, \alpha_n) = L(T\alpha_1, \ldots, T\alpha_n)$$

define uma forma L_T, n-linear alternada. Seja M um gerador do módulo $\Lambda^n(V)$ de posto 1. Cada L em $\Lambda^n(V)$ pode ser escrito de modo único como $L = aM$ para algum a em K. Em particular, $M_T = cM$ para um determinado c. Para $L = aM$, temos

$$\begin{aligned} L_T &= (AM)_T \\ &= aM_T \\ &= a(cM) \\ &= c(aM) \\ &= cL. \end{aligned}$$

É claro que o elemento c, neste último corolário, é denominado o determinante de T. De (5-39), para o caso $r = n$ (quando existe somente uma r-upla crescente $J = (1, \ldots, n)$), vemos que o determinante de T é o determinante da matriz que representa T em qualquer base ordenada $\{\beta_1, \ldots, \beta_n\}$. Vejamos por que. O elemento i, j da matriz representante é

$$A_{ij} = f_j(T\beta_i)$$

de modo que

$$D_J(T\beta_1, \ldots, T\beta_n) = \sum_\sigma (\text{sinal } \sigma) A(1, \sigma 1) \ldots A(n, \sigma n)$$
$$= \det A.$$

Por outro lado,

$$D_J(T\beta_1, \ldots, T\beta_n) = (\det T)\, D_J(\beta_1, \ldots, \beta_n)$$
$$= \det T.$$

A razão destas observações é que, por intermédio do Teorema 7 e seu corolário, obtemos uma definição do determinante de um operador linear, que não requer o conhecimento de determinantes de matrizes. Determinantes de matrizes podem ser definidos em termos de determinantes de operadores ao contrário do que fizemos na Seção 5.4.

Queremos dizer algo mais sobre as formas especiais D_J, r-lineares alternadas, associadas a uma base $\{f_1, \ldots, f_n\}$ de V^* em (5-39). É importante entender que $D_J(\alpha_1, \ldots, \alpha_r)$ é o determinante de uma determinada $r \times r$ matriz. Se

$$A_{ij} = f_j(\alpha_i), \qquad 1 \leq i \leq r,\ 1 \leq j \leq n,$$

isto é, se

$$\alpha_i = A_{i1}\beta_1 + \ldots + A_{in}\beta_n, \qquad 1 \leq i \leq r$$

e J é a r-upla crescente $(j_1, \ldots, j_r)$, então

(5-42) $$D_J(\alpha_1, \ldots, \alpha_r) = \sum_\sigma (\text{sinal } \sigma) A(1, j_{\sigma 1}) \ldots A(n, j_{\sigma n})$$
$$= \det \begin{bmatrix} A(1, j_1) & \ldots & A(1, j_r) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A(r, j_1) & \ldots & A(r, j_r) \end{bmatrix}$$

Assim, $D_J(\alpha_1, \ldots, \alpha_r)$ é o determinante da $r \times r$ matriz, formada pelas colunas $j_1, \ldots, j_r$ da $r \times n$ matriz, que tem para linhas (as n-uplas coordenadas de) $\alpha_1, \ldots, \alpha_r$. Uma outra notação, algumas vezes usada para este determinante, é

(5-43) $$D_J(\alpha_1, \ldots, \alpha_r) = \frac{\partial(\alpha_1, \ldots, \alpha_r)}{\partial(\beta_{j1}, \ldots, \beta_{jr})}$$

Nesta notação, a demonstração do Teorema 7 mostra que toda forma r-linear, alternada L pode ser expressa em relação a uma base $\{\beta_1, \ldots, \beta_n\}$ pela equação

$$L(\alpha_1, \ldots, \alpha_r) = \sum_{j_1 < \ldots < j_r} \frac{\partial(\alpha_1, \ldots, \alpha_r)}{\partial(\beta_{j_1}, \ldots, \beta_{j_r})} L(\beta_{j_1}, \ldots, \beta_{j_r}). \tag{5-44}$$

5.7 O Anel de Grassman

Muitas das propriedades importantes de determinantes e formas multilineares alternadas podem ser descritas em termos de uma operação de multiplicação sobre formas, denominada produto exterior. Se L e M são formas sobre o módulo V, respectivamente r e s-lineares, a elas está associado um produto de L e M, o produto tensorial $L \otimes M$. Ele não define uma forma alternada a não ser quando $L = 0$ ou $M = 0$; no entanto, existe uma maneira natural de projetá-lo em $\Lambda^{r+s}(V)$. Aparentemente

$$L \cdot M = \pi_{r+s}(L \otimes M) \tag{5-45}$$

deveria ser a multiplicação "natural" de formas alternadas. Mas será que é?

Tomemos um exemplo específico. Suponhamos que V seja o módulo K^n e $f_1, \ldots, f_n$ as funções coordenadas canônicas sobre K^n. Se $i \neq j$,

$$f_i \cdot f_j = \pi_2(f_i \otimes f_j)$$

é a função (determinante)

$$D_{ij} = f_i \otimes f_j - f_j \otimes f_i$$

dada por (5-39). Suponhamos agora que k seja um índice diferente de i e de j. Então

$$\begin{aligned} D_{ij} \cdot f_k &= \pi_3[(f_i \otimes f_j - f_j \otimes f_i) \otimes f_k] \\ &= \pi_3(f_i \otimes f_j \otimes f_k) - \pi_3(f_j \otimes f_i \otimes f_k). \end{aligned}$$

A demonstração do Lema, que está após a equação (5-36), mostra que para uma forma L, r-linear, arbitrária e para qualquer permutação σ de $\{1, \ldots, r\}$

$$\pi_r(L_\sigma) = \text{sinal } \sigma \ \pi_r(L).$$

Logo, $D_{ij} \cdot f_k = 2\pi_3(f_i \otimes f_j \otimes f_k)$. Por um cálculo análogo $f_i \cdot D_{jk} = 2\pi_3(f_i \otimes f_j \otimes f_k)$. Portanto, temos

$$(f_i \cdot f_j) \cdot f_k = f_i \cdot (f_j \cdot f_k)$$

e tudo isto parece muito promissor. Mas há um problema. Apesar do cálculo que acabamos de efetuar, a multiplicação admitida em (5-45) não é associativa. De fato, se l for um índice diferente de i, j, k, os cálculos mostram que

$$D_{ij} \cdot D_{kl} = 4\pi_4(f_i \otimes f_j \otimes f_k \otimes f_l)$$

e

$$(D_{ij} \cdot f_k) \cdot f_l = 6\pi_4(f_i \otimes f_j \otimes f_k \otimes f_l).$$

Assim, em geral

$$(f_i \cdot f_j) \cdot (f_k \cdot f_l) \neq [(f_i \cdot f_j) \cdot f_k] \cdot f_l$$

e vemos que nossa primeira tentativa de achar uma multiplicação, produziu uma operação não associativa.

O leitor não deve ficar surpreso com a monotonia de uma verificação direta das duas equações que mostram a não associatividade. Isto é típico do assunto, como também é típico que exista um fato geral que simplifique consideravelmente o trabalho.

Suponhamos que L seja uma forma r-linear e M uma forma s-linear sobre o módulo V. Então

$$\begin{aligned}\pi_{r+s}((\pi_r L) \otimes (\pi_s M)) &= \pi_{r+s}\Big(\sum_{\sigma,\tau} (\text{sinal } \sigma)(\text{sinal } \tau)\, L_\sigma \otimes M_\tau\Big)\\ &= \sum_{\sigma,\tau} (\text{sinal } \sigma)(\text{sinal } \tau)\, \pi_{r+s}(L_\sigma \otimes M_\tau)\end{aligned}$$

com σ variando sobre o grupo simétrico S_r das permutações de $\{1, \ldots, r\}$ e τ variando sobre S_s. Cada par σ, τ define um elemento (σ, τ) de S_{r+s} que permuta os primeiros r elementos de $\{1, \ldots, r+s\}$ de acordo com σ e os últimos s elementos de acordo com τ. É claro que

$$\text{sinal}(\sigma,\tau) = (\text{sinal } \sigma)(\text{sinal } \tau)$$

e que

$$(L \otimes M)_{(\sigma,\tau)} = L_\sigma \otimes L_\tau.$$

Portanto,

$$\pi_{r+s}[(\pi_r L) \otimes (\pi_s M)] = \sum_{\sigma,\tau} \text{sinal}(\sigma, \tau)\pi_{r+s}[(L \otimes M)_{(\sigma,\tau)}].$$

Ora, já observamos que

$$\text{sinal}\,(\sigma, \tau)\pi_{r+s}[(L \otimes M)_{(\sigma,\tau)}] = \pi_{r+s}(L \otimes M).$$

Assim, segue-se que

(5-46) $$\pi_{r+s}[(\pi_r L) \otimes (\pi_s M)] = r!s!\, \pi_{r+s}(L \otimes M).$$

Esta fórmula simplifica numerosos cálculos. Por exemplo, suponhamos que $I = (i_1, \ldots, i_r)$ seja uma r-upla crescente e que $J = (j_1, \ldots, j_s)$ seja uma s-upla crescente. Para simplificar as coisas, suponhamos ainda que

$$i_1 < \ldots < i_r < j_1 < \ldots < j_s.$$

Então temos as funções determinantes associadas

$$D_I = \pi_r(E_I)$$
$$D_J = \pi_s(E_J)$$

onde E_I e E_J são dadas por (5-30). Usando (5-46), vemos imediatamente que

$$\begin{aligned} D_I \cdot D_J &= \pi_{r+s}[\pi_r(E_I) \otimes \pi_s(E_J)] \\ &= r!s!\pi_{r+s}(E_I \otimes E_J). \end{aligned}$$

Como $E_I \otimes E_J = E_{I \cup J}$, segue-se que

$$D_I \cdot D_J = r!s!D_{I \cup J}.$$

Isto sugere que a ausência de associatividade para a multiplicação (5-45) decorre do fato de que $D_I \cdot D_J \neq D_{I \cup J}$. Afinal, o produto de D_I e D_J deveria ser $D_{I \cup J}$. Para consertar a situação, deveríamos definir um novo produto, o **produto exterior** de uma forma r-linear alternada L e uma forma s-linear alternada M por

(5-47) $$L \wedge M = \frac{1}{r!s!}\pi_{r+s}(L \otimes M).$$

Temos então

$$D_I \wedge D_J = D_{I \cup J}$$

para as funções determinantes sobre K^n, e se existir alguma justiça neste mundo, deveríamos ter encontrado uma boa multiplicação para formas multilineares alternadas. Infelizmente, (5-47) não tem sentido para o caso mais geral sob consideração, pois talvez não seja possível dividir por $r!s!$ no anel K. Se K for um corpo de característica zero, então (5-47) tem sentido e pode-se rapidamente demonstrar que o produto exterior é associativo.

Teorema 8. *Seja* K *um corpo de característica zero e* V *um espaço vetorial sobre* K. *Então o produto exterior é uma operação associativa sobre as formas multilineares alternadas sobre* V. *Em outras palavras, se* L, M *e* N *são formas multilineares alternadas de graus* r, s *e* t, *respectivamente, então*

$$(\mathrm{L} \wedge \mathrm{M}) \wedge \mathrm{N} = \mathrm{L} \wedge (\mathrm{M} \wedge \mathrm{N}).$$

Demonstração. Decorre de (5-47) que $cd(L \wedge M) = cL \wedge dM$ para escalares c e d arbitrários. Logo

$$r!s!t![(L \wedge M) \wedge N] = r!s!(L \wedge M) \wedge t!N$$

e como $\pi_t(N) = t!N$, resulta que

$$\begin{aligned} r!s!t![(L \wedge M) \wedge N] &= \pi_{r+s}(L \otimes M) \wedge \pi_t(N) \\ &= \frac{1}{(r+s)!}\frac{1}{t!}\pi_{r+s+t}[\pi_{r+s}(L \otimes M) \otimes \pi_t(N)]. \end{aligned}$$

De (5-46) vemos agora que

$$r!s!t[(L \wedge M) \wedge N] = \pi_{r+s+t}(L \otimes M \otimes N).$$

Por um cálculo análogo

$$r!s!t![L \wedge (M \wedge N)] = \pi_{r+s+t}(L \otimes M \otimes N)$$

e, portanto,

$$(L \wedge M) \wedge N = L \wedge (M \wedge N).$$

Voltemos agora ao caso geral, onde somente suporemos que K seja um anel comutativo com elemento unidade. Nosso primeiro problema é substituir (5-47) por uma definição equivalente e que funcione em geral. Se L e M forem formas multilineares alternadas de graus r e s, respectivamente, construiremos uma forma multilinear $L \wedge M$, alternada, canônica, de grau $r+s$, tal que

$$r!s!(L \wedge M) = \pi_{r+s}(L \otimes M).$$

Lembremos a definição de $\pi_{r+s}(L \otimes M)$. A cada permutação σ de $\{1, \ldots, r+s\}$ associamos a função multilinear

(5-48) $$(\text{sinal } \sigma)\ (L \otimes M)_\sigma$$

onde

$$(L \otimes M)_\sigma(\alpha_1, \ldots, \alpha_{r+s}) = (L \otimes M)\,(\alpha_{\sigma 1}, \ldots, \alpha_{\sigma(r+s)})$$

e somamos as funções (5-48) sobre todas as permutações σ. Existem $(r+s)!$ permutações; no entanto, como L e M são alternadas, muitas das funções (5-48) são iguais. De fato, existem no máximo

$$\frac{(r+s)!}{r!s!}$$

funções distintas (5-48). Vejamos por que. Seja S_{r+s} o conjunto das permutações de $\{1, \ldots, r+s\}$, isto é, seja S_{r+s} o grupo simétrico de grau $r+s$. Como na demonstração de (5-46), distinguiremos o subconjunto G das permutações σ que permutam entre si os elementos de $\{1, \ldots, r\}$ e $\{r+1, \ldots, r+s\}$. Em outras palavras, σ está em G se $1 \leq \sigma i \leq r$ para cada i entre 1 e r. (Decorre necessariamente que $r+1 \leq \sigma j \leq r+s$ para cada j entre $r+1$ e $r+s$). Ora, G é um subgrupo de S_{r+s}, isto é, se σ e τ estiverem em G, $\sigma\tau^{-1}$ estará em G. Evidentemente G possui $r!s!$ elementos.

Temos uma aplicação

$$S_{r+s} \xrightarrow{\psi} M^{r+s}(V)$$

definida por

$$\psi(\sigma) = (\text{sinal } \sigma)\ (L \otimes M)_\sigma.$$

Como L e M são alternadas

$$\psi(\gamma) = L \otimes M$$

para todo γ em G. Portanto, como $(N\sigma)\tau = N\tau\sigma$ para toda forma $(r+s)$-linear N sobre V, temos

$$\psi(\tau\gamma) = \psi(\tau), \qquad \tau \text{ em } S_{r+s}, \ \gamma \text{ em } G.$$

Isto mostra que a aplicação ψ é constante sobre cada **classe lateral** (à esquerda) τG do subgrupo G. Se τ_1 e τ_2 estão em S_{r+s}, as classes laterais $\tau_1 G$ e $\tau_2 G$ são idênticas ou disjuntas conforme $\tau_2^{-1}\tau_1$ esteja ou não em G. Cada classe lateral contém $r!s!$ elementos; logo, existem

$$\frac{(r+s)!}{r!s!}$$

classes laterais distintas. Se S_{r+s}/G indicar a coleção de classes laterais, então ψ define uma função sobre S_{r+s}/G, isto é, pelo que mostramos, existe uma função $\tilde{\psi}$ sobre este conjunto, tal que

$$\psi(\tau) = \tilde{\psi}(\tau G)$$

para todo τ em S_{r+s}. Se H é uma classe lateral à esquerda de G, então $\tilde{\psi}(H) = \psi(\tau)$ para todo τ em H.

Definamos agora o **produto exterior** das formas multilineares alternadas L e S, de graus r e s, colocando

$$L \wedge M = \sum_{H} \tilde{\psi}(H) \tag{5-49}$$

onde H varia sobre S_{r+s}/G. Uma outra maneira de enunciar a definição de $L \wedge M$ é a seguinte: seja S um conjunto arbitrário de permutações de $\{1, \ldots, r+s\}$ que contenha exatamente um elemento de cada classe lateral à esquerda de G. Então

$$L \wedge M = \sum_{\sigma} (\text{sinal } \sigma)\, (L \otimes M)_\sigma \tag{5-50}$$

onde σ varia sobre S. Evidentemente

$$r!s!L \wedge M = \pi_{r+s}(L \otimes M)$$

de modo que a nova definição é equivalente a (5-47) quando K for um corpo de característica zero.

Teorema 9. *Seja* K *um anel comutativo com elemento unidade e seja* **V** *um módulo sobre* K. *Então o produto exterior é uma operação associativa sobre as formas multilineares alternadas sobre* **V**. *Em outras palavras, se* L, M *e* N *são formas multilineares alternadas sobre* **V** *de graus* r, s *e* t, *respectivamente, então*

$$(L \wedge M) \wedge N = L \wedge (M \wedge N).$$

Demonstração. Apesar de não podermos aplicar aqui a demonstração do Teorema 8, ela sugere como devemos tratar o caso geral. Seja (r, s, t) o subgrupo de S_{r+s+t} constituído das permutações que permutam entre si os elementos dos conjuntos

$$\{1, \ldots, t\}, \{r+1, \ldots, r+s\}, \{r+s+1, \ldots, r+s+t\}.$$

Então (sinal μ) $(L \otimes M \otimes N)_\mu$ é a mesma função multilinear para todos μ numa dada classe lateral à esquerda de $G(r, s, t)$. Escolhamos um elemento de cada classe lateral à esquerda de $G(r, s, t)$ e seja E a soma dos termos correspondentes (sinal μ) $(L \otimes M \otimes N)_\mu$. Então E é independente da maneira em que os representantes μ são escolhidos, e

$$r!s!t!E = \pi_{r+s+t}(L \otimes M \otimes N).$$

Devemos mostrar que $(L \wedge M) \wedge N$ e $L \wedge (M \wedge N)$ são ambos iguais a E.

Seja $G(r+s, t)$ o subgrupo de S_{r+s+t} que permuta entre si os elementos dos conjuntos

$$\{1, \ldots, r+s\}, \{r+s+1, \ldots, r+s+t\}.$$

Seja T uma permutação arbitrária de $\{1, \ldots, r+s+t\}$ que contenha exatamente um elemento de cada classe lateral à esquerda de $G(r+s, t)$. De (5-50)

$$(L \wedge M) \wedge N = \sum_\tau (\text{sinal } \tau) [(L \wedge M) \otimes N]_\tau$$

onde a somatória se estende sobre as permutações τ em T. Ora, seja $G(r, s)$ o subgrupo de S_{r+s} que permuta entre si os elementos dos conjuntos

$$\{1, \ldots, r\}, \{r+1, \ldots, r+s\}.$$

Seja S um conjunto arbitrário de permutações de $\{1, \ldots, r+s\}$ que contenha exatamente um elemento de cada classe lateral à esquerda de $G(r, s)$. De (5-50) e do que foi visto acima, decorre que

$$(L \wedge M) \wedge N = \sum_{\sigma,\tau} (\text{sinal } \sigma)(\text{sinal } \tau) \, [(L \otimes M \otimes N)_\sigma]_\tau$$

onde a somatória se estende sobre todos os pares σ, τ em $S \times T$. Se convencionarmos identificar cada σ em S_{r+s} com o elemento de S_{r+s+t} que age como σ sobre $\{1, \ldots, r+s\}$ e é igual à função idêntica sobre $\{r+s+1, \ldots, r+s+t\}$, podemos escrever

$$(L \wedge M) \wedge N = \sum_{\sigma,\tau} \text{sinal } (\sigma \tau) \, [(L \otimes M \otimes N_\sigma]_\tau.$$

Mas

$$[(L \otimes M \otimes N)_\sigma]_\tau = (L \otimes M \otimes N)_{\tau\sigma}.$$

Portanto,

$$(L \wedge M) \wedge N = \sum_{\sigma,\tau} \text{sinal } (\tau \sigma) \, (L \otimes M \otimes N)_{\tau\sigma}.$$

Suponhamos agora que

$$\tau_1 \sigma_1 = \tau_2 \sigma_2 \gamma$$

com σ_i em S, τ_i em T e γ em $G(r, s, t)$. Então $\tau_2^{-1} \tau_1 = \sigma_2 \gamma \sigma_1^{-1}$, e como $\sigma_2 \gamma \sigma_1^{-1}$ está em $G(r+s, t)$, segue-se que τ_1 e τ_2 estão na mesma classe lateral à esquerda de $G(r+s, t)$. Portanto, $\tau_1 = \tau_2$ e $\sigma_1 = \sigma_2 \gamma$. Mas isto implica que σ_1 e σ_2 (considerados como elementos de S_{r+s}) estão na mesma classe lateral de $G(r, s)$; logo $\sigma_1 = \sigma_2$. Portanto, os produtos $\tau\sigma$ correspondentes aos

$$\frac{(r+s+t)!}{(r+s)!\,t!} \, \frac{(r+s)!}{r!\,s!}$$

pares (τ, σ) em $T \times S$ são todos distintos e estão em classes laterais distintas de $G(r, s, t)$. Como existem em S_{r+s+t}, exatamente

$$\frac{(r+s+t)!}{r!\,s!\,t!}$$

classes laterais à esquerda de $G(r, s, t)$ segue-se que $(L \wedge M) \wedge N = E$. Por um argumento análogo, também $L \wedge (M \wedge N) = E$.

Exemplo 13. O produto exterior está intimamente relacionado com certas fórmulas usadas para calcular determinantes e conhecidas sob o nome de **desenvolvimento segundo Laplace.** Seja K um anel comutativo com elemento unidade e seja n um inteiro positivo. Suponhamos que $1 \leq r < n$ e seja L a forma r-linear alternada sobre K^n, definida por

$$L(\alpha_1, \ldots, \alpha_r) = \det \begin{bmatrix} A_{11} & \cdots & A_{1r} \\ \vdots & & \vdots \\ A_{r1} & \cdots & A_{rr} \end{bmatrix}.$$

Se $s = n - r$ e M é a forma s-linear alternada

$$M(\alpha_1, \ldots, \alpha_s) = \det \begin{bmatrix} A_{1(r+1)} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{s(r+1)} & & A_{sn} \end{bmatrix}$$

então $L \wedge M = D$, a função determinante sobre K^n. Isto decorre imediatamente do fato de que $L \wedge M$ é uma forma n-linear alternada e (como pode ser visto)

$$(L \wedge M)(\varepsilon_1, \ldots, \varepsilon_n) = 1.$$

Se agora descrevermos $L \wedge M$ da maneira correta, obteremos um desenvolvimento segundo Laplace do determinante de uma $n \times n$ matriz sobre K.

No grupo das permutações S_n, seja G o subgrupo que permuta entre si os elementos dos conjuntos $\{1, \ldots, r\}$ e $\{r+1, \ldots, n\}$. Cada classe lateral à esquerda de G contém precisamente uma permutação σ tal que $\sigma 1 < \sigma 2 < \ldots < \sigma r$ e $\sigma(r+1) < \ldots < \sigma n$. O sinal desta permutação é dado por

$$\text{sinal } \sigma = (-1)^{\sigma 1 + \cdots + \sigma r + (r(r-1)/2)}$$

O produto exterior $L \wedge M$ é dado por

$$(L \wedge M)(\alpha_1, \ldots, \alpha_n) = \Sigma\, (\text{sinal } \sigma)\, L(\alpha\sigma_1, \ldots, \alpha_{\sigma r})\, M(\alpha_{\sigma(r+1)}, \ldots, \alpha_{\sigma n})$$

onde a somatória se estende sobre uma coleção de σ, um de cada classe lateral de G. Portanto,

$$(L \wedge M)(\alpha_1, \ldots, \alpha_n) = \sum_{j_1 < \cdots < j_r} e_J \, L(\alpha_{j_1}, \ldots, \alpha_{j_r}) \, M(\alpha_{k_1}, \ldots, \alpha_{k_s})$$

onde

$$e_J = (-1)^{j_1 + \ldots + j_r + /r(r-1)/2)}$$
$$k_i = \sigma(r + i).$$

Em outras palavras,

$$\det A = \sum_{j_1 < \ldots < j_r} e_J \begin{vmatrix} A_{j_1,1} & \cdots & A_{j_1,r} \\ \vdots & & \vdots \\ A_{j_r,1} & \cdots & A_{j_r,r} \end{vmatrix} \begin{vmatrix} A_{k_1,r+1} & \cdots & A_{k_1,n} \\ \vdots & & \vdots \\ A_{k_r,r+1} & \cdots & A_{k_r,n} \end{vmatrix}$$

Este é um desenvolvimento segundo Laplace. Outros podem ser obtidos, substituindo os conjuntos $\{1, \ldots, r\}$ e $\{r+1, \ldots, n\}$ por outros dois conjuntos complementares de índices.

Se V é um K-módulo, podemos reunir os diversos módulos $\Lambda^r(V)$ e usar o produto exterior para definir um anel. Para simplificar, faremos isto apenas para o caso de um K-módulo livre de posto n. Os módulos $\Lambda^r(V)$ são então triviais para $r > n$. Definamos

$$\Lambda(V) = \Lambda^0(V) \oplus \Lambda^1(V) \oplus \ldots \oplus \Lambda^n(V).$$

Esta é uma soma direta externa — algo ainda não discutido. Os elementos de $\Lambda(V)$ são as $(n+1)$-uplas $(L_0, \ldots, L_n)$ com L_r em $\Lambda^r(V)$. A adição e a multiplicação por elementos de K são definidas da maneira esperada para $(n+1)$-uplas. Aliás, $\Lambda^0(V) = K$. Se identificarmos $\Lambda^r(K)$ com as $(n+1)$-uplas $(0, \ldots, 0, L, 0, \ldots, 0)$ com L em $\Lambda^r(K)$, então $\Lambda^r(K)$ é um submódulo de $\Lambda(V)$ e a decomposição em soma direta

$$\Lambda(V) = \Lambda^0(V) \oplus \ldots \oplus \Lambda^n(V)$$

é válida no sentido usual. Como $\Lambda^r(V)$ é um K-módulo livre de posto $\binom{n}{r}$, vemos que $\Lambda(V)$ é um K-módulo livre e

$$\text{posto } \Lambda(V) = \sum_{r=0}^{n} \binom{n}{r}$$
$$= 2^n.$$

O produto exterior define uma multiplicação em $\Lambda(V)$; basta usar o produto exterior sobre formas e estendê-lo linearmente para $\Lambda(V)$. Ele é distributivo sobre a adição em $\Lambda(V)$ e define em $\Lambda(V)$ uma estrutura de anel. Este anel é o **anel de Grassman** sobre V^*. Ele não é comutativo, pois, por exemplo, se L e M estiverem, respectivamente em Λ^r e Λ^s, então

$$L \wedge M = (-1)^{rs} M \wedge L.$$

Mas, o anel de Grassman é importante em várias partes da matemática.

Capítulo 6
FORMAS CANÔNICAS ELEMENTARES

6.1 Introdução

Mencionamos anteriormente que nossa meta principal é estudar transformações lineares sobre espaços vetoriais de dimensão finita. Até este ponto já vimos muitos exemplos particulares de transformações lineares e demonstramos alguns teoremas sobre transformações lineares arbitrárias. No caso de dimensão finita utilizamos bases ordenadas para representar essas transformações por meio de matrizes e essa representação nos ajuda a perceber o seu comportamento. Pesquisamos o espaço vetorial $L(V, W)$ das transformações lineares de um espaço em outro e também a álgebra linear $L(V, V)$ das transformações lineares de um espaço em si mesmo.

Nos próximos dois capítulos, estaremos preocupados com operadores lineares. Nosso programa é tomar um operador linear T sobre um espaço vetorial V de dimensão finita e "desmontá-lo para ver como ele funciona". De início, será mais fácil exprimir nosso objetivo em linguagem matricial: Dado um operador linear T, determinar uma base ordenada para V, em relação à qual a matriz de T assuma uma forma especialmente simples.

Eis uma ilustração do que temos em mente. Com exceção dos múltiplos escalares da matriz unidade, as matrizes mais simples talvez sejam as matrizes diagonais:

(6-1)
$$D = \begin{bmatrix} c_1 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ 0 & 0 & c_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & c_n \end{bmatrix}$$

Seja T um operador linear sobre um espaço V, n-dimensional. Se conseguíssemos determinar uma base ordenada $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V, em relação à qual T fosse representado por uma matriz diagonal D (6-1), obteríamos bastante informações sobre T. Por exemplo, números simples associados a T, como o posto de T ou o determinante de T, poderiam ser determinados com nada mais que um simples exame da matriz D. Poderíamos descrever explicitamente a imagem e o núcleo de T. Como $[T]_{\mathscr{B}} = D$ se, e somente se,

$$T\alpha_k = c_k\alpha_k \qquad k = 1, \ldots, n \tag{6-2}$$

a imagem será o subespaço gerado por aqueles α_k para os quais $c_k \neq 0$ e o núcleo será gerado pelos α_k restantes. De fato, parece ser justo dizer, que se conhecêssemos uma base $\mathscr{B}$ e uma matriz diagonal D tal que $[T]_{\mathscr{B}} = D$, poderíamos responder imediatamente a qualquer pergunta que pudesse surgir a respeito de T.

Será que qualquer operador T pode ser representado por uma matriz diagonal em relação a alguma base ordenada? Em caso negativo, para quais operadores T existe uma tal base? Se uma tal base existir, como poderemos determiná-la? E se não existir, qual é o tipo mais simples de matriz que pode representar T? Estas são algumas das questões que trataremos neste (e no próximo) capítulo. À medida que aprendermos quais são as dificuldades, a forma das nossas questões tornar-se-á mais sofisticada.

6.2 Valores Característicos

As observações da introdução, na seção anterior, nos fornecem um ponto de partida para a nossa tentativa de analisar um operador linear arbitrário T. A expressão (6-2) sugere que estudemos vetores que sejam levados por T em múltiplos escalares de si mesmos.

Definição. *Seja* V *um espaço vetorial sobre o corpo* F *e seja* T *um operador linear sobre* V. *Um* **valor característico** *de* T *é um escalar* c *em* F *tal que exista um vetor não-nulo* α *em* V *com* $T\alpha = c\alpha$. *Se* c *é um valor característico de* T, *então*

(a) *todo* α *tal que* $T\alpha = c\alpha$ *é chamado* **vetor característico** *de* T *associado ao valor característico* c;

(b) *a coleção de todos os* α *tais que* $T\alpha = c\alpha$ *é denominado* **espaço característico** *associado a* c.

Valores característicos são freqüentemente denominados raízes características, raízes latentes, autovalores, valores próprios ou valores espectrais. Neste livro usaremos apenas o nome "valores característicos".

Se T é um operador linear arbitrário e c um escalar qualquer, o conjunto dos vetores α tais que $T\alpha = c\alpha$ é um subespaço de V. Esse subespaço é o núcleo da transformação linear $(T - cI)$. Dizemos que c é um valor característico de T se esse subespaço for diferente do subespaço nulo, isto é, se $(T - cI)$ não for injetora. Se o espaço subjacente V for de dimensão finita, $(T - cI)$ deixará de ser injetora precisamente quando o seu determinante for diferente de zero. Em resumo:

Teorema 1. *Seja* T *um operador linear sobre um espaço* V *de dimensão finita e seja* c *um escalar. As seguintes condições são equivalentes:*

(i) c *é um valor característico de* T.

(ii) *O operador* (T − cI) *é singular* (*não-inversível*).

(iii) *det*(T − cI) = 0.

O critério (iii) é muito importante, pois nos diz onde procurar os valores característicos de T. Como $\det(T - cI)$ é um polinômio de grau n na variável c, os valores característicos serão as raízes desse polinômio. Expliquemos com mais cuidado.

Se $\mathscr{B}$ é uma base ordenada arbitrária de V e $A = [T]_{\mathscr{B}}$, então $(T - cI)$ é inversível se, e somente se, a matriz $(A - cI)$ é inversível. Portanto, colocamos a definição que segue.

Definição. *Se* A *é uma* n × n *matriz sobre o corpo* F, *um* **valor característico** *de* A *em* F *é um escalar* c *em* F *tal que a matriz* (A − cI) *seja singular* (*não-inversível*).

Como c é um valor característico de A se, e somente se, $\det(A - cI) = 0$, ou equivalentemente, se, e somente se, $\det(cI - A) = 0$, podemos formar a matriz $(xI - A)$ cujos elementos são polinômios e considerar o polinômio $f = \det(xI - A)$. Evidentemente os valores característicos de A em F são exatamente os escalares c em F tais que $f(c) = 0$. Por esta razão f é denominado o **polinômio característico** de A. É importante notar que f é um polinômio unitário cujo grau é exatamente n. Isto pode ser visto facilmente através da fórmula para o determinante de uma matriz em função de seus elementos.

Lema. *Matrizes semelhantes têm o mesmo polinômio característico.*

Demonstração. Se $B = P^{-1}AP$, então

$$\begin{aligned} \det(xI - B) &= \det(xI - P^{-1}AP) \\ &= \det(P^{-1}(xI - A)P) \\ &= \det P^{-1} \cdot \det(xI - A) \cdot \det P \\ &= \det(xI - A). \end{aligned}$$

Este lema nos permite definir sensatamente o polinômio característico do operador T como o polinômio característico de qualquer $n \times n$ matriz que represente T em relação a alguma base ordenada de V. Da mesma forma que para matrizes, os valores característicos de T serão as raízes do polinômio característico de T. Em particular, isto nos mostra que T não pode ter mais que n valores característicos distintos. É importante ressaltar que T pode não ter nenhum valor característico.

Exemplo 1. Seja T o operador linear sobre R^2 que é representado em relação à base ordenada canônica pela matriz.

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

O polinômio característico de T (ou de A) é

$$\det(xI - A) = \begin{vmatrix} x & 1 \\ -1 & x \end{vmatrix} = x^2 + 1.$$

Como este polinômio não possui raízes reais, T não possui valores característicos. Se U é o operador linear sobre C^2 que é representado por A em relação à base ordenada canônica, então U possui dois valores característicos, i e $-i$. Vemos aqui um ponto sutil. Ao discutirmos os valores característicos de uma matriz A, precisamos tomar o cuidado de estipular o corpo envolvido. A matriz A acima não possui nenhum valor característico em $\mathbb{R}$, mas possui dois valores característicos, i e $-i$, em C.

Exemplo 2. Seja A a 3×3 matriz (real)

$$\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}.$$

O polinômio característico de A é

$$\begin{vmatrix} x-3 & -1 & 1 \\ -2 & x-2 & 1 \\ -2 & -2 & x \end{vmatrix} = x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2.$$

Logo, os valores característicos de A são 1 e 2.

Suponhamos que T seja o operador linear sobre R^3, representado por A em relação à base canônica. Determinemos os vetores característicos de T, associados aos valores característicos 1 e 2. Ora,

$$A - I = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix}.$$

Obviamente, o posto de $A - I$ é 2 (e portanto a nulidade de $T - I$ é 1). Assim, o espaço dos vetores característicos associados ao valor característico 1, tem dimensão 1. O vetor $\alpha_1 = (1, 0, 2)$ gera o núcleo de $T - I$. Logo, $T\alpha = \alpha$ se, e somente se, α é um múltiplo escalar de α_1. Consideremos agora

$$A - 2I = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix}.$$

Evidentemente, o posto de $A - 2I$ também é 2, de modo que o espaço dos vetores característicos associado ao valor característico 2 tem dimensão 1. É evidente que $T\alpha = 2\alpha$ se, e somente se, α é um múltiplo escalar de $\alpha_2 = (1, 1, 2)$.

Definição. *Seja* T *um operador linear sobre o espaço* V *de dimensão finita. Dizemos que* T *é* **diagonalizável** *se existe uma base de* V *formada por vetores característicos de* T.

A razão para o nome deveria ser evidente; de fato, se existe uma base ordenada $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V na qual cada α_i é um vetor característico de T, então a matriz T em relação à base ordenada $\mathscr{B}$ é diagonal. Se $T\alpha_i = c_i\alpha_i$, então

$$[T]_{\mathscr{B}} = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix}.$$

Certamente não pretendemos que os escalares $c_1, \ldots, c_n$ sejam distintos; na verdade, eles podem ser todos iguais (quando T é um múltiplo escalar do operador idêntico).

Poder-se-ia também definir T como sendo diagonalizável quando os vetores característicos de T gerassem V. Isto difere apenas superficialmente de nossa definição, uma vez que podemos sempre obter uma base a partir de qualquer conjunto de vetores que seja gerador.

Para os exemplos 1 e 2 escolhemos propositalmente operadores lineares T sobre R^n não diagonalizáveis. No Exemplo 1 temos um operador linear sobre R^2 que não é diagonalizável pois não possui valores característicos. No Exemplo 2, o operador T tem valores característicos; de fato, o polinômio característico de T é completamente fatorável sobre o corpo dos números reais: $f = (x-1)(x-2)^2$. Apesar disso T não é diagonalizável. Existe somente um espaço de dimensão um, de vetores característicos, associado a cada um dos dois valores característicos de T. Portanto, não há possibilidade de formar uma base de R^3 constituída de vetores característicos de T.

Suponhamos que T seja um operador linear diagonalizável. Sejam $c_1, \ldots, c_k$ os valores característicos *distintos* de T. Então existe uma base ordenada $\mathscr{B}$ em relação à qual T é representado por uma matriz diagonal cujos elementos diagonais são os escalares c_i, cada um repetido um determinado número de vezes. Se c_i está repetido d_i vezes, então (podemos fazer com que) a matriz tenha a forma em blocos

$$(6\text{-}3) \qquad [T]_{\mathscr{B}} = \begin{bmatrix} c_1 I_1 & 0 & \cdots & 0 \\ 0 & c_2 I_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_k I_k \end{bmatrix}$$

onde I_j é a $d_j \times d_j$ matriz unidade. A partir dessa matriz verificamos dois fatos. Primeiro, o polinômio característico de T é o produto de fatores lineares (possivelmente repetidos):

$$f = (x - c_1)^{d_1} \ldots (x - c_k)^{d_k}.$$

Se o corpo F dos escalares for algebricamente fechado, como por exemplo o corpo dos números complexos, todo polinômio sobre F poderá ser fatorado dessa maneira (ver Seção 4.5); no entanto,

se F não for algebricamente fechado, estaremos citando uma propriedade especial de T, ao dizermos que seu polinômio característico tem uma tal fatoração. O segundo fato que observamos a partir de (6-3) é que d_i, o número de vezes que c_i é repetido como raiz de f, é igual à dimensão do espaço dos vetores característicos associado ao valor característico c_k. Isto acontece porque a nulidade da matriz é igual ao número de zeros que ela possui na sua diagonal principal, e a matriz $[T - c_i I]_{\mathscr{B}}$ tem d_i zeros na sua diagonal principal. A relação entre a dimensão do espaço característico e a multiplicidade do valor característico como uma raiz de f, não parece ser muito emocionante à primeira vista; no entanto ela nos fornecerá uma maneira mais simples de verificar se um dado operador é ou não diagonalizável.

Lema. *Suponhamos que* $T\alpha = c\alpha$. *Se* f *é um polinômio arbitrário, então* $f(T)\alpha = f(c)\alpha$.

Demonstração. Exercício.

Lema. *Seja* T *um operador linear sobre o espaço* V *de dimensão finita. Sejam* $c_1, \ldots, c_k$ *valores característicos distintos de* T *e seja* W_i *o espaço dos vetores característicos associado ao valor característico* c_i. *Se* $W = W_1 + \ldots + W_k$, *então*

$$\dim W = \dim W_1 + \ldots + \dim W_k.$$

De fato, se $\mathscr{B}_i$ *é uma base ordenada de* W_i, *então* $\mathscr{B} = (\mathscr{B}_1, \ldots, \mathscr{B}_k)$ *é uma base ordenada de* W.

Demonstração. O espaço $W = W_1 + \ldots + W_k$ é o subespaço gerado por todos os vetores característicos de T. Normalmente, quando formamos a soma W de subespaços W_i, esperamos que $\dim W < \dim W_1 + \ldots + \dim W_k$ por causa das relações lineares que possam existir entre os vetores nos diversos espaços. Esse lema afirma que os espaços característicos associados a valores característicos distintos, são independentes um do outro.

Suponhamos que (para cada i) β_i seja um vetor em W_i e que $\beta_1 + \ldots + \beta_k = 0$. Mostraremos que $\beta_i = 0$ para cada i. Seja f um polinômio arbitrário. Como $T\beta_i = c_i\beta_i$, o lema anterior nos diz que

$$\begin{aligned} 0 = f(T)0 &= f(T)\beta_1 + \ldots + f(T)\beta_k \\ &= f(c_1)\beta_1 + \ldots + f(c_k)\beta_k \end{aligned}$$

Escolhamos polinômios $f_1, \ldots, f_k$ tais que

$$f_i(c_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Então

$$0 = f_i(T)0 = \sum_j \delta_{ij}\beta_j$$
$$= \beta_2.$$

Seja, agora, $\mathscr{B}_i$ uma base ordenada de W_i e seja $\mathscr{B}$ a seqüência $\mathscr{B} = (\mathscr{B}_1, \ldots, \mathscr{B}_k)$. Então $\mathscr{B}$ gera o subespaço $W = W_1 + \ldots + W_k$. Também, $\mathscr{B}$ é uma seqüência de vetores linearmente independentes, pelo seguinte: qualquer relação linear entre os vetores de $\mathscr{B}$ terá a forma $\beta_1 + \ldots + \beta_k = 0$, onde β_i é alguma combinação linear dos vetores de $\mathscr{B}_i$. Pelo que acabamos de ver, sabemos que $\beta_i = 0$ para cada i. Como cada $\mathscr{B}_i$ é linearmente independente, vemos que existe apenas a relação linear trivial entre os vetores em $\mathscr{B}$.

Teorema 2. *Seja* T *um operador linear sobre um espaço vetorial* V *de dimensão finita. Sejam* $c_1, \ldots, c_k$ *valores característicos distintos de* T *e seja* W_i *o núcleo de* $(T - c_iI)$. *As seguintes afirmações são equivalentes*:

(i) T *é diagonalizável.*

(ii) *O polinômio característico de* T *é*

$$f = (x - c_1)^{d_1} \ldots (x - c_k)^{d_k}$$

e dim $W_i = d_i$, $i = 1, \ldots, k$.

(iii) *dim* $W_1 + \ldots +$ *dim* $W_k =$ *dim* V.

Demonstração. Já observamos que (i) implica (ii). Se o polinômio característico f for o produto de fatores lineares como em (ii), então $d_1 + \ldots + d_k = \dim V$. De fato, a soma dos d_i é o grau do polinômio característico e este grau é dim V. Portanto, (ii) implica (iii). Suponhamos (iii). Pelo lema, devemos ter $V = W_1 + \ldots + W_k$, isto é, os vetores característicos de T geram V.

O análogo do Teorema 2, usando matrizes, pode ser formulado da seguinte maneira: Seja A uma $n \times n$ matriz com elementos em um corpo F e sejam $c_1, \ldots, c_k$ os valores característicos distintos de A em F. Para cada i, seja W_i o espaço das matrizes coluna X (com elementos de F) tais que

$$(A - c_i I)X = 0,$$

e seja $\mathscr{B}_i$ uma base ordenada de W_i. As bases $\mathscr{B}_1, \ldots, \mathscr{B}_k$ podem ser reunidas para formar a seqüência de colunas de uma matriz P:

$$P = [P_1, P_2, \ldots] = (\mathscr{B}_1, \ldots, \mathscr{B}_k).$$

A matriz A é semelhante sobre F a uma matriz diagonal se, e somente se, P é uma matriz quadrada. Quando P é quadrada, P inversível e $P^{-1}AP$ é diagonal.

Exemplo 3. Seja T o operador linear sobre R^3 representado em relação à base ordenada canônica pela matriz

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}.$$

Mostraremos como se pode calcular o polinômio característico por meio de diversas operações sobre linhas e colunas:

$$\begin{vmatrix} x-5 & 6 & 6 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{vmatrix} = \begin{vmatrix} x-5 & 0 & 6 \\ 1 & x-2 & -2 \\ -3 & 2-x & x+4 \end{vmatrix}$$

$$= (x-2)\begin{vmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -3 & -1 & x+4 \end{vmatrix}$$

$$= (x-2)\begin{vmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -2 & 0 & x+2 \end{vmatrix}$$

$$= (x-2)\begin{vmatrix} x-5 & 6 \\ -2 & x+2 \end{vmatrix}$$

$$= (x-2)(x^2 - 3x + 2)$$

$$= (x-2)^2(x-1).$$

Quais são as dimensões dos espaços dos vetores característicos associados aos dois valores característicos? Temos

$$A - I = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix}$$

$$A - 2I = \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix}.$$

Sabemos que $A - I$ é singular e obviamente posto $(A - I) \geq 2$. Portanto, posto$(A - I) = 2$. É evidente que posto $(A - 2I) = 1$.

Sejam W_1 e W_2 os espaços dos vetores característicos associados, respectivamente, aos valores característicos 1 e 2. Sabemos que dim $W_1 = 1$ e dim $W_2 = 2$. Pelo Teorema 2, T é diagonalizável. É fácil construir uma base de R^3 em relação à qual T seja representado por uma matriz diagonal. O núcleo de $(T - I)$ é gerado pelo vetor $\alpha_1 = (3, -1, 3)$ e assim $\{\alpha_1\}$ é uma base de W_1. O núcleo de $T - 2I$ (isto é, o espaço W_2) é constituído pelos vetores (x_1, x_2, x_3) com $x_1 = 2x_2 + 2x_3$. Assim, um exemplo de uma base de W_2 é $\{\alpha_2, \alpha_3\}$, com

$$\begin{aligned} \alpha_2 &= (2, 1, 0) \\ \alpha_3 &= (2, 0, 1). \end{aligned}$$

Se $\mathscr{B} = \{\alpha_1, \alpha_2, \alpha_3\}$, então $[T]_{\mathscr{B}}$, é a matriz diagonal

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

O fato de que T é diagonalizável significa que a matriz original A é semelhante (sobre R) à matriz diagonal D. A matriz P que nos possibilita mudar coordenadas da base $\mathscr{B}$ para a base canônica é (naturalmente) a matriz que tem como vetores colunas as transpostas de $\alpha_1, \alpha_2, \alpha_3$:

$$P = \begin{bmatrix} 3 & 2 & 2 \\ -1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}.$$

Além do mais, $AP = PD$, de modo que

$$P^{-1}AP = D.$$

Exercícios

1. Em cada um dos casos seguintes, seja T o operador linear sobre R^2 que é representado pela matriz A em relação à base ordenada canônica de R^2 e seja U o operador linear sobre C^2 representado por A em relação à base ordenada canônica. Determinar os polinômios característicos de T e de U, determinar os valores característicos de cada operador e para cada valor característico c determinar uma base do correspondente espaço de vetores característicos.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

2. Seja V um espaço vetorial n-dimensional sobre F. Qual é o polinômio característico do operador idêntico sobre V? Qual é o polinômio característico do operador nulo?

3. Seja A uma $n \times n$ matriz triangular sobre o corpo F. Demonstrar que os valores característicos de A são os elementos diagonais de A, isto é, os escalares A_{ii}.

4. Seja T o operador linear sobre R^3 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}.$$

Demonstrar que T é diagonalizável mostrando uma base de R^3, formada por vetores característicos de T.

5. Seja

$$A = \begin{bmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{bmatrix}$$

A é semelhante sobre o corpo R a alguma matriz diagonal? Esta matriz é semelhante sobre o corpo C a alguma matriz diagonal?

6. Seja T o operador linear sobre R^4 representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{bmatrix}$$

Em que condições sobre a, b e c, T é diagonalizável?

7. Seja T um operador linear sobre o espaço vetorial n-dimensional V e suponhamos que T possua n valores característicos *distintos*. Demonstrar que T é diagonalizável.

8. Seja A e B $n \times n$ matrizes sobre o corpo F. Demonstrar que se $(I - AB)$ é inversível, então $I - BA$ é inversível e

$$(I - BA)^{-1} = I + B(I - AB)^{-1}A.$$

9. Usar o resultado do Exercício 8 para demonstrar que se A e B são $n \times n$ matrizes sobre o corpo F, então AB e BA têm exatamente os mesmos valores característicos em F.

10. Suponhamos que A seja uma 2×2 matriz com elementos reais que seja simétrica $(A^t = A)$. Demonstrar que A é semelhante sobre R a uma matriz diagonal.

11. Seja N uma 2×2 matriz complexa tal que $N^2 = 0$. Demonstrar que ou $N = 0$ ou N é semelhante sobre C a

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

12. Usar o resultado do Exercício 1 para demonstrar o seguinte: Se A é uma 2×2 matriz com elementos complexos, então A é semelhante sobre C a uma matriz de um dos tipos seguintes:

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \quad \begin{bmatrix} a & 0 \\ 1 & a \end{bmatrix}.$$

13. Seja V o espaço vetorial de todas as funções f de R em R que são contínuas, isto é, o espaço das funções contínuas, definidas sobre a reta real e tomando valores reais. Seja T o operador linear sobre V definido por

$$(Tf)(x) = \int_0^x f(t)\, dt.$$

Demonstrar que T não possui nenhum valor característico.

14. Seja A uma $n \times n$ matriz *diagonal* com polinômio característico

$$(x - c_1)^{d_1} \dots (x - c_k)^{d_k}$$

sendo $c_1, \dots, c_k$ distintos. Seja V o espaço das $n \times n$ matrizes B tais que $AB = BA$. Demonstrar que a dimensão de V é $d_1^2 + \dots + d_k^2$.

15. Seja V o espaço das $n \times n$ matrizes sobre F. Seja A uma $n \times n$ matriz fixa sobre F. Seja T o operador linear "multiplicação à esquerda por A" sobre V. É verdade que A e T possuem os mesmos valores característicos?

6.3 Polinômios Anuladores

Ao tentarmos analisar um operador linear T, é de grande utilidade conhecermos a classe dos polinômios que anulam T. Especificamente, suponhamos que T seja um operador linear sobre V, espaço vetorial sobre o corpo F. Se p for um polinômio sobre F, então $p(T)$ será novamente um operador linear sobre V. Se q for um outro polinômio sobre F, então

$$(p+q)(T) = p(T) + q(T)$$
$$(pq)(T) = p(T)\,q(T).$$

Portanto, a coleção de polinômios p que anulam T, no sentido de que

$$p(T) = 0,$$

é um ideal da álgebra dos polinômios $F[x]$. Poderia ser o ideal nulo, isto é, poderia ser que T não fosse anulado por nenhum polinômio não-nulo. Mas, isso não poderá acontecer se o espaço V for de dimensão finita.

Suponhamos que T seja um operador linear sobre o espaço n-dimensional V. Observemos as primeiras (n^2+1) potências de T:

$$I, T, T^2, \ldots, T^{n^2}.$$

Essa é uma seqüência de n^2+1 operadores em $L(V, V)$, o espaço dos operadores lineares sobre V. O espaço $L(V, V)$ tem dimensão n^2. Portanto, essa seqüência de n^2+1 operadores é necessariamente linearmente dependente, isto é, temos

$$c_0 I + c_1 T + \ldots + c_{n^2} T^{n^2} = 0$$

para escalares c_i, não todos nulos. Portanto, o ideal dos polinômios que anulam T contém um polinômio não-nulo de grau menor ou igual a n^2.

De acordo com o Teorema 5 do Capítulo 4, todo ideal de polinômios consiste de todos os múltiplos de algum polinômio unitário fixo, o gerador do ideal. Portanto, podemos associar ao operador T um polinômio unitário p com essa propriedade: Se f é um polinômio sobre F, então $f(T) = 0$ se, e somente se, $f = pg$, onde g é algum polinômio sobre F.

Definição. *Seja* T *um operador linear sobre um espaço vetorial* V *de dimensão finita sobre o corpo* F. *O* **polinômio minimal** *de* T *é o (único) gerador unitário do ideal dos polinômios sobre* F *que anulam* T.

O nome "polinômio minimal" deriva do fato de que o gerador de um ideal de polinômios é o polinômio unitário de menor grau no ideal. Isso significa que o polinômio minimal p do operador linear T é determinado de modo único pelas tres propriedades seguintes:

(1) p é um polinômio unitário sobre o corpo F dos escalares.

(2) $p(T) = 0$.

(3) Nenhum polinômio sobre F que anule T tem grau menor que o de p.

Se A é uma $n \times n$ matriz sobre F, definimos o polinômio minimal de A, de maneira análoga, como sendo o único gerador unitário do ideal formado pelos polinômios sobre F que anulam A. Se o operador T for representado, em relação a alguma base ordenada, pela matriz A, então T e A terão o mesmo polinômio minimal. Isso acontece porque $f(T)$ é representado, em relação a esta base, pela matriz $f(A)$, de modo que $f(T) = 0$ se, e somente se, $f(A) = 0$.

Dessas últimas observações sobre operadores e matrizes decorre que matrizes semelhantes possuem o mesmo polinômio minimal. Este fato decorre também das definições, pois

$$f(P^{-1}AP) = P^{-1}f(A)P$$

para todo polinômio f.

Existe uma outra observação fundamental que devemos fazer a respeito de polinômios minimais de matrizes. Suponhamos que A seja uma $n \times n$ matriz com elementos em um corpo F. Suponhamos que F_1 seja um corpo contendo F como subcorpo. (Por exemplo, A poderia ser uma matriz com elementos racionais, enquanto que F_1 seria o corpo dos números reais. Ou, A seria uma matriz com elementos reais e F_1 o corpo dos números complexos). Podemos considerar A como uma $n \times n$ matriz sobre F ou como uma $n \times n$ matriz sobre F_1. À primeira vista, poderia parecer que obteríamos dois polinômios minimais distintos para A. Felizmente isso não acontece e devemos ver por que. Qual é a definição de polinômio minimal de A, quando A é considerada como uma $n \times n$ matriz sobre o corpo F? Consideremos todos os polinômios

unitários com coeficientes em F que anulem A, e escolhamos aquele de menor grau. Se f é um polinômio unitário sobre F:

(6-4) $$f = x^k + \sum_{j=0}^{k-1} a_j x^j$$

$f(A) = 0$ apenas afirma que existe uma relação linear entre as potências de A:

(6-5) $$A^k + a_{k-1}A^{k-1} + \ldots + a_1 A + a_0 I = 0.$$

O grau do polinômio minimal é o menor inteiro positivo k tal que exista uma relação linear da forma (6-5) entre as potências I, $A, \ldots, A^k$. Além disso, pela unicidade do polinômio minimal, existe para aquele k uma, e uma só, relação da forma (6-5); isto é, uma vez determinado o k mínimo, existem escalares $a_0, \ldots, a_{k-1}$, determinados de modo único em F, tais que (6-5) seja válido. Estes escalares são os coeficientes do polinômio minimal.

Ora (para cada k) temos em (6-5) um sistema de n^2 equações lineares nas "incógnitas" $a_0, \ldots, a_{k-1}$. Como os coeficientes de A pertencem a F, os coeficientes do sistema de equações (6-5) estarão em F. Portanto, se o sistema possuir uma solução com $a_0, \ldots, a_{k-1}$ em F_1, ele possuirá uma solução com $a_0, \ldots, a_{k-1}$ em F. (Veja o fim da Seção 1.4.) Deveria estar claro agora que os dois polinômios minimais são iguais.

O que sabemos, até aqui, a respeito de polinômio minimal de um operador linear sobre um espaço n-dimensional? Apenas que o seu grau não excede n^2. Esta é uma estimativa muito pobre pois o grau não pode exceder n. Demonstraremos mais adiante que o operador é anulado pelo seu polinômio característico. Primeiramente, porém, observemos um fato mais elementar.

Teorema 3. *Seja* T *um operador linear sobre um espaço vetorial* n-*dimensional* V [*ou, seja* A *uma* n × n *matriz*]. *Os polinômios característico e minimal de* T (*de* A) *possuem as mesmas raízes, a menos de multiplicidades.*

Demonstração. Seja p o polinômio minimal de T. Seja c um escalar. Queremos mostrar que $p(c) = 0$ se, e somente se, c é um valor característico de T.

Primeiramente, suponhamos $p(c) = 0$. Então

$$p = (x - c)q$$

onde q é um polinômio. Como $\text{gr}(q) < \text{gr}(p)$, a definição de polinômio minimal p nos diz que $q(T) \neq 0$. Escolhamos um vetor β tal que $q(T)\beta \neq 0$. Seja $\alpha = q(T)\beta$. Então

$$\begin{aligned} 0 &= p(T)\beta \\ &= (T - cI)q(T)\beta \\ &= (T - cI)\alpha \end{aligned}$$

e assim, c é um valor característico de T.

Suponhamos, agora, que c seja um valor característico de T, digamos, $T\alpha = c\alpha$ com $\alpha \neq 0$. Como observamos no lema anterior,

$$p(T)\alpha = p(c)\alpha.$$

Como $p(T) = 0$ e $\alpha \neq 0$, temos $p(c) = 0$.

Seja T um operador linear diagonalizável e sejam $c_1, \ldots, c_k$ os valores distintos de T. É fácil ver que o polinômio minimal de T é o polinômio

$$p = (x - c_1) \ldots (x - c_k).$$

Se α é um vetor característico, então um dos operadores $T - c_1 I$, $\ldots$, $T - c_k I$ leva α em 0. Portanto,

$$(T - c_1 I) \ldots (T - c_k I)\alpha = 0$$

para todo vetor característico α. Existe uma base para o espaço subjacente constituída de vetores característicos de T; logo

$$p(T) = (T - c_1 I) \ldots (T - c_k I) = 0.$$

O que concluímos é o seguinte: Se T é um operador linear diagonalizável, o polinômio minimal de T é um produto de fatores lineares distintos. Como veremos mais adiante, esta propriedade caracteriza os operadores diagonalizáveis.

Exemplo 4. Tentemos determinar os polinômios mínimais dos operadores dos exemplos 1, 2 e 3. Vamos discuti-los em ordem inversa. Vimos que o operador do Exemplo 3 é diagonalizável e que seu polinômio característico é

$$f = (x - 1)(x - 2)^2.$$

Sabemos, pelo parágrafo precedente, que o polinômio minimal de T é

$$p = (x - 1)(x - 2).$$

O leitor talvez ache confortante verificar diretamente que

$$(A - I)(A - 2I) = 0.$$

No Exemplo 2, o polinômio característico do operador T também é $f = (x - 1)(x - 2)^2$. Mas este T não é diagonalizável e portanto não sabemos se o polinômio minimal é $(x - 1)(x - 2)$. O que sabemos a respeito do polinômio minimal neste caso? Usando o Teorema 3, sabemos que suas raízes são 1 e 2, sendo permitidas algumas multiplicidades. Portanto, procuremos p entre polinômios da forma $(x - 1)^k(x - 2)^l$, $k \geq 1$, $l \geq 1$. Tentemos $(x - 1)(x - 2)$:

$$(A - I)(A - 2I) = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix}$$
$$= \begin{bmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{bmatrix}.$$

Assim, o grau do polinômio minimal é pelo menos 3. Devemos então tentar $(x - 1)^2(x - 2)$ ou $(x - 1)(x - 2)^2$. Esse último, sendo o polinômio característico, pareceria ser uma escolha menos ao acaso. É fácil verificar que $(A - I)(A - 2I)^2 = 0$. Portanto, o polinômio minimal de T é o seu polinômio característico.

No Exemplo 1 discutimos o operador linear T sobre R^2 representado, em relação à base canônica, pela matriz

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

O seu polinômio característico é $x^2 + 1$ e não possui raízes reais. Para determinar o polinômio minimal, esqueçamos T e concentremo-nos em A. Como uma 2×2 matriz complexa, A tem valores característicos i e $-i$. Ambas as raízes são necessariamente raízes do polinômio minimal. Portanto, o polinômio minimal é divisível por $x^2 \times 1$. É trivial verificar que $A^2 + I = 0$. Portanto, o polinômio minimal é $x^2 + 1$.

Teorema 4 (Cayley-Hamilton). *Seja* T *um operador linear sobre um espaço vetorial* V *de dimensão finita. Se* f *é o polinômio característico de* T, *então* f(T) = 0; *em outras palavras, o polinômio minimal divide o polinômio característico de* T.

Demonstração. Mais tarde daremos duas demonstrações deste resultado, independentes da demonstração que será dada aqui. Esta demonstração, apesar de curta, talvez seja difícil de compreender. Além da brevidade, ela tem a virtude de oferecer uma aplicação ilustrativa e não trivial da teoria geral dos determinantes, desenvolvida no Capítulo 5.

Seja K o anel comutativo com elemento unidade constituído por todos os polinômios em T. É claro que K é realmente uma álgebra comutativa com elemento unidade sobre o corpo dos escalares. Escolhamos uma base ordenada $\{\alpha_1, \ldots, \alpha_n\}$ de V e seja A a matriz que representa T em relação à base dada. Então

$$T\alpha_i = \sum_{j=1}^{n} A_{ji}\alpha_j, \qquad 1 \leq j \leq n.$$

Essas equações podem ser escritas na forma equivalente

$$\sum_{j=1}^{n} (\delta_{Ij}T - A_{ji}I)\alpha_j = 0, \qquad 1 \leq i \leq n.$$

Indiquemos por B o elemento de $K^{n \times n}$ com argumentos

$$B_{ij} = \delta_{ij}T - A_{ji}I.$$

Quando $n = 2$

$$B = \begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix}$$

e

$$\begin{aligned} \det B &= (T - A_{11}I)(T - A_{22}I) - A_{12}A_{21}I \\ &= T^2 - (A_{11} + A_{22})T + (A_{11}A_{22} - A_{12}A_{21})I \\ &= f(T) \end{aligned}$$

onde f é o polinômio característico:

$$f = x^2 - (\text{traço } A)x + \det A.$$

Para o caso $n > 2$, é claro que

$$\det B = f(T)$$

pois f é o determinante da matriz $xI - A$ cujos elementos são os polinômios

$$(xI - A)_{ij} = \delta_{ij}x - A_{ji}.$$

Queremos mostrar que $f(T) = 0$. Para que $f(T)$ seja o operador nulo é necessário e suficiente que $(\det B)\alpha_k = 0$ para $k = 1, \ldots, n$. Pela definição de B, os vetores $\alpha_1, \ldots, \alpha_n$ satisfazem as equações

$$\sum_{j=1}^{n} B_{ij}\alpha_j = 0, \qquad 1 \leq i \leq n. \tag{6-6}$$

Quando $n = 2$, é sugestivo escrever (6-6) na forma

$$\begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Nesse caso, a adjunta clássica, adj B é a matriz

$$\tilde{B} = \begin{bmatrix} T - A_{22}I & A_{21}I \\ A_{12}I & T - A_{11}I \end{bmatrix}$$

e

$$\tilde{B}B = \begin{bmatrix} \det B & 0 \\ 0 & \det B \end{bmatrix}.$$

Assim, temos

$$\begin{aligned} (\det B)\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} &= (\tilde{B}B)\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \\ &= \tilde{B}\left(B\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

No caso geral, seja $\tilde{B} = \text{adj } B$. Então, por (6-6)

$$\sum_{j=1}^{n} \tilde{B}_{ki}B_{ij}\alpha_j = 0$$

para cada par k, i, e somando em relação a i, temos

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{B}_{ki}B_{ij}\alpha_j \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \tilde{B}_{ki}B_{ij} \right) \alpha_j. \end{aligned}$$

Ora, $\tilde{B}B = (\det \tilde{B})I$, de modo que

$$\sum_{i=1}^{n} \tilde{B}_{ki}B_{ij} = \delta_{kj} \det B.$$

Portanto,

$$\begin{aligned} 0 &= \sum_{j=1}^{n} \delta_{kj}(\det B)\alpha_j \\ &= (\det B)\alpha_k, \qquad 1 \leq k \leq n. \end{aligned}$$

O teorema de Cayley-Hamilton é útil para nós, neste ponto, principalmente porque abrevia a pesquisa do polinômio minimal de diversos operadores. Se conhecermos a matriz A que representa T em relação a alguma base ordenada, podemos determinar o polinômio característico f. Sabemos que o polinômio minimal p divide f e que os dois polinômios possuem as mesmas raízes. Não existe nenhum método que permita determinar precisamente as raízes de um polinômio (a não ser que seu grau seja pequeno); no entanto, se f é fatorável

(6-7) $\quad f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}, \qquad c_1, \dots, c_k$ distintos, $d_i \geq 1$

então

(6-8) $\quad p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}, \qquad 1 \leq r_j \leq d_j.$

Em geral, é só isso que podemos dizer. Se f é o polinômio (6-7) e tiver grau n, então para todo polinômio p, como em (6-8), pode-

mos determinar uma $n \times n$ matriz que tenha polinômio característico f e polinômio minimal p. Não vamos demonstrar esse fato agora. Mas queremos ressaltar que o fato de sabermos que o polinômio característico tem a forma (6-7) nos diz que o polinômio minimal tem a forma (6-8) e nada mais nos diz a respeito de p.

Exemplo 5. Seja A a 4×4 matriz (racional)

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

É fácil calcular as potências de A:

$$A^2 = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix}. \qquad A^3 = \begin{bmatrix} 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \end{bmatrix}.$$

Assim, $A^3 = 4A$, isto é, se $p = x^3 - 4x = x(x+2)(x-2)$, $p(A) = 0$. O polinômio minimal de A necessariamente divide p. Obviamente o grau do polinômio minimal não é 1 pois isto significaria que A é um múltiplo escalar da matriz unidade. Portanto, os candidatos para polinômio minimal são. $p, x(x+2)$, $x(x-2)$, $x^2 - 4$. Os polinômios quadráticos podem ser eliminados pois é óbvio que $A^2 \neq -2A$, $A^2 \neq 2A$ e $A^2 \neq 4I$. Portanto, p é o polinômio minimal de A. Em particular, 0, 2 e -2 são os valores característicos de A. Um dos fatores x, $x-2$ ou $x+2$, deve estar repetido duas vezes no polinômio característico. Evidentemente, $\text{posto}(A) = 2$. Portanto, existe um espaço bi-dimensional de vetores característicos associado ao valor característico 0. Usando o Teorema 2, deve ser evidente, agora, que o polinômio característico é $x^2(x^2-4)$ e que A é semelhante, sobre o corpo dos números racionais, à matriz

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}.$$

Exercícios

1. Seja V um espaço vetorial de dimensão finita. Qual é o polinômio minimal do operador idêntico sobre V? Qual é o polinômio minimal do operador nulo?

2. Sejam a, b e c elementos de um corpo F e seja A a seguinte 3×3 matriz sobre F:

$$\begin{bmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{bmatrix}$$

Demonstrar que o polinômio característico de A é $x^3 - ax^2 - bx - c$, e que este polinômio é também o polinômio minimal de A.

3. Seja A a 4×4 matriz real

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -2 & 2 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix}$$

Demonstrar que o polinômio característico de A é $x^2(x-1)^2$ e que esse polinômio é também o polinômio minimal.

4. A matriz A do exercício 3 é semelhante sobre o corpo dos números complexos a alguma matriz diagonal?

5. Seja V um espaço vetorial n-dimensional e seja T um operador linear sobre V. Suponhamos que exista um inteiro positivo k tal que $T^k = 0$. Demonstrar que $T^n = 0$.

6. Determinar uma 3×3 matriz cujo polinômio minimal seja x^2.

7. Seja n um inteiro positivo e seja V o espaço dos polinômios sobre R de grau no máximo igual a n (acrescentar o polinômio nulo). Seja D o operador derivação sobre V. Qual é o polinômio minimal de D?

8. Seja P o operador sobre R^2 que projeta cada vetor sobre o eixo-x, paralelamente ao eixo-y: $P(x, y) = (x, 0)$. Mostrar que P é linear. Qual é o polinômio minimal de P?

9. Seja A uma $n \times n$ matriz com polinômio característico

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}.$$

Demonstrar que

$$c_1 d_1 + \dots + c_k d_k = \text{traço}(A).$$

10. Seja V o espaço vetorial das $n \times n$ matrizes sobre o corpo F. Seja A uma $n \times n$ matriz fixa. Seja T o operador linear definido por

$$T(B) = AB.$$

Demonstrar que o polinômio minimal de T é o polinômio minimal de A.

11. Sejam A e B $n \times n$ matrizes sobre o corpo F. De acordo com o exercício 9 da seção 6.1, as matrizes AB e BA possuem os mesmos valores característicos. Elas possuem o mesmo polinômio característico? Elas possuem o mesmo polinômio minimal?

6.4 Subespaços Invariantes

Nessa seção serão introduzidos alguns conceitos úteis na tentativa de analisar um operador linear. Usaremos essas idéias para obter caracterizações de operadores diagonalizáveis (e trianguláveis) em termos de seus polinômios minimais.

Definição. *Seja* V *um espaço vetorial e* T *um operador linear sobre* V. *Se* W *é um subespaço de* V, *dizemos que* W *é* **invariante sob** T *se para cada vetor* α *em* W, *o vetor* Tα *está em* W, *isto é, se* T(W) *está contido em* W.

Exemplo 6. Se T é um operador linear arbitrário sobre V, então V é invariante sob T, da mesma forma que o é o subespaço nulo. A imagem de T e o núcleo de T também são invariantes sob T.

Exemplo 7. Seja F um corpo e seja D o operador derivação sobre o espaço $F[x]$ dos polinômios sobre F. Seja n um inteiro positivo e seja W o subespaço dos polinômios de grau menor ou igual a n. Então W é invariante sob D. Esta é apenas outra maneira de dizer que D diminui o grau.

Exemplo 8. Aqui está uma generalização muito útil do Exemplo 6. Seja T um operador linear sobre V. Seja U um operador linear arbitrário sobre V que comuta com T, isto é, $TU = UT$. Sejam W a imagem de U e N o núcleo de U. Tanto W como N são invariantes sob T. Se α está na imagem de U, digamos $\alpha = U\beta$, então $T\alpha = T(U\beta) = U(T\beta)$, de modo que $T\alpha$ estará na imagem de U. Se α está em N, então $U(T\alpha) = T(U\alpha) = T(0) = 0$; logo, $T\alpha$ estará em N.

Um tipo particular de operador que comuta com T é um operador $U = g(T)$ onde g é um polinômio. Por exemplo, podemos ter $U = T - cI$, onde c é o valor característico de T. Conhecemos o núcleo de U. Vemos que este exemplo inclui o fato (óbvio) de

que o espaço dos vetores característicos de T associado ao valor característico c, é invariante sob T.

Exemplo 9. Seja T o operador linear sobre R^2 que é representado em relação à base ordenada canônica pela matriz

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Então os únicos subespaços de R^2 que são invariantes sob T são R^2 e o subespaço nulo. Qualquer outro subespaço invariante teria necessariamente dimensão 1. Mas se W é o subespaço gerado por algum vetor não-nulo α, a afirmação de que W é invariante sob T significa que α é um vetor característico, mas A não possui valores característicos reais.

Quando o subespaço W é invariante sob o operador T, então T induz um operador linear T_W sobre o espaço W. O operador linear T_W é definido por $T_W(\alpha) = T(\alpha)$, para α em W, mas T_W é um objeto bem diferente de T uma vez que seu domínio é W e não V.

Quando V é de dimensão finita, a invariância de W sob T admite uma interpretação simples por meio de matrizes e talvez devamos mencioná-la neste ponto. Suponhamos que tomemos uma base ordenada $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V tal que $\mathscr{B}' = \{\alpha_1, \ldots, \alpha_r\}$ seja uma base ordenada de $W (r = \dim W)$. Seja $A = [T]_{\mathscr{B}}$. Então

$$T\alpha_j = \sum_{i=1}^{n} A_{ij}\alpha_i.$$

Como W é invariante sob T, o vetor $T\alpha_j$ pertence a W para $j \leq r$. Isto significa que

$$T\alpha_j = \sum_{i=1}^{r} A_{ij}\alpha_i, \quad j \leq r. \tag{6-9}$$

Em outras palavras, $A_{ij} = 0$ se $j \leq r$ e $i > r$.

Esquematicamente, A é da forma em blocos

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \tag{6-10}$$

onde B é uma $r \times r$ matriz, C é uma $r \times (n-r)$ matriz e D é uma $(n-r) \times (n-r)$ matriz. O leitor deverá notar que, de acordo com (6-9), a matriz B é exatamente a matriz do operador induzido T_W, em relação à base ordenada $\mathscr{B}'$.

Na maioria das vezes desenvolveremos argumentos sobre T e T_W sem fazer uso da forma em blocos da matriz A em (6-10). Devemos observar, porém, como certas relações entre T_W e T tornam-se aparentes nessa forma em blocos.

Lema. *Seja* W *um subespaço invariante sob* T. *O polinômio característico do operador* T_W *(restrição de* T *a* W) *divide o polinômio característico de* T. *O polinômio minimal de* T_W *divide o polinômio minimal de* T.

Demonstração. Temos

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

onde $A = [T]_{\mathscr{B}}$ e $B = [T_W]$. Por causa da forma em blocos dessa matriz

$$\det(xI - A) = \det(xI - B)\det(xI - D).$$

Isso demonstra a afirmação quanto aos polinômios característicos. Observemos que o símbolo I foi usado para representar matrizes unidades de três tipos diferentes.

A k-ésima potência da matriz A é da forma em blocos

$$A^k = \begin{bmatrix} B^k & C_k \\ 0 & D^k \end{bmatrix}.$$

onde C_k é uma $r \times (n-r)$ matriz. Portanto, qualquer polinômio que anula A também anula B (e D também). Assim, o polinômio minimal de B divide o polinômio minimal de A.

Exemplo 10. Seja T um operador linear arbitrário sobre um espaço V de dimensão finita. Seja W o subespaço gerado por *todos* os vetores característicos de T. Sejam $c_1, \ldots, c_k$ valores característicos distintos de T. Para cada i, seja W_i o espaço dos vetores característicos associados ao valor característico c_i e seja $\mathscr{B}_i$ uma

base ordenada de W_i. O lema que precede o Teorema 2 nos diz que $\mathcal{B}' = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ é uma base ordenada de W. Em particular,

$$\dim W = \dim W_1 + \ldots + \dim W_k.$$

Seja $\mathcal{B}' = \{\alpha_1, \ldots, \alpha_r\}$ tal que os primeiros α formam a base $\mathcal{B}_1$, os seguintes $\mathcal{B}_2$ e assim por diante. Então

$$T\alpha_i = t_i\alpha_i, \qquad i = 1, \ldots, r$$

onde $(t_1, \ldots, t_r) = (c_1, c_1, \ldots, c_1, \ldots, c_k, c_k, \ldots, c_k)$ com c_i repetido $\dim W_i$ vezes.

Ora, W é invariante sob T, pois para cada α em W temos

$$\begin{aligned} \alpha &= x_1\alpha_1 + \ldots + x_r\alpha_r \\ T\alpha &= t_1x_1\alpha_1 + \ldots + t_rx_r\alpha_r \end{aligned}$$

Escolhamos quaisquer outros vetores $\alpha_{r+1}, \ldots, \alpha_n$ em V tais que $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ seja uma base de V. A matriz de T em relação à $\mathcal{B}$ é da forma em blocos (6-10) e a matriz do operador T_W, restrição de T a W, em relação à base $\mathcal{B}'$ é

$$B = \begin{bmatrix} t_1 & 0 & \ldots & 0 \\ 0 & t_2 & \ldots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \ldots & t_r \end{bmatrix}.$$

O polinômio característico de B (isto é, de T_W) é

$$g = (x - c_1)^{e_1} \ldots (x - c_k)^{e_k}$$

com $e_i = \dim W_i$. Além do mais, g divide f, o polinômio característico de T. Portanto, a multiplicidade de c_i, como uma raiz de f, é no mínimo $\dim W_i$.

Tudo isso deveria deixar o Teorema 2 absolutamente claro. Ele simplesmente afirma que T é diagonalizável se, e somente se, $r = n$, isto é, se, e somente se, $e_1 + \ldots + e_k = n$. Ele não é de grande ajuda nos casos não diagonalizáveis, pois não conhecemos as matrizes C e D de (6-10).

Definição. *Seja* W *um subespaço invariante sob* T *e seja* α *um vetor de* V. *O* **T-condutor de** α **em** W *é o conjunto* $S_T(\alpha; W)$ *constituído de todos os polinômios* g *(sobre o corpo dos escalares) tais que* g(T)α *esteja em* W.

Como o operador T permanecerá fixo durante a maioria das discussões, abandonaremos, em geral, o índice T e escreveremos $S(\alpha; W)$. Muitos autores chamam esta coleção de polinômios de "estofo" ("*das einstopfende Ideal*"). "Condutor" é o termo padrão, preferido por aqueles que visualizam um operador $g(T)$ menos agressivo, gentilmente conduzindo o vetor α para dentro de W. No caso especial em que $W = \{0\}$, o condutor é denominado o T-**anulador de** α.

Lema. *Se* W *é um subespaço invariante de* T, *então* W *é invariante sob qualquer polinômio em* T. *Assim, para cada* α *em* V, *o condutor* S(α; W) *é um ideal na álgebra dos polinômios* F[x].

Demonstração. Se β está em W, então $T\beta$ estará em W. Conseqüentemente, $T(T\beta) = T^2\beta$ estará em W. Por indução, $T^k\beta$ estará em W, para cada k. Tomamos combinações lineares, para concluir que $f(T)\beta$ está em W para todo polinômio f.

A definição de $S(\alpha; W)$ tem sentido mesmo quando W é um subconjunto arbitrário de V. Se W é um subespaço, então $S(\alpha; W)$ é um subespaço de $F[x]$ porque

$$(cf + g)(T) = cf(T) + g(T).$$

Suponhamos, além disso, que W seja invariante sob T e seja g um polinômio em $S(\alpha; W)$, isto é, seja $g(T)\alpha$ um elemento de W. Se f é um polinômio arbitrário, então $f(T)[g(T)\alpha]$ está em W. Como

$$(fg)(T) = f(T)g(T)$$

fg está em $S(\alpha; W)$. Portanto, o condutor absorve o produto por qualquer polinômio.

O único gerador unitário do ideal $S(\alpha; W)$ é também denominado T-**condutor de** α em W (o T-**anulador** no caso em que $W = \{0\}$). O T-condutor de α em W é o polinômio unitário g de menor grau tal que $g(T)\alpha$ esteja em W. Um polinômio f está em $S(\alpha; W)$ se, e somente se, g divide f. Observemos que o condutor $S(\alpha; W)$ contém sempre o polinômio minimal de T; logo, *todo T-condutor divide o polinômio minimal de T.*

Como primeira ilustração de como usar o condutor $S(\alpha; W)$ caracterizaremos os operadores trianguláveis. O operador linear T se diz **triangulável** se existir uma base ordenada em relação à qual T seja representado por uma matriz triangular.

Lema. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F. *Seja* T *um operador linear sobre* V *tal que o polinômio minimal de* T *seja um produto de fatores lineares*

$$p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}, \qquad c_i \text{ em } F.$$

Seja W *um subespaço próprio de* $V (W \neq V)$ *invariante sob* T. *Então existe um vetor* α *em* V *tal que*

(a) α *não está em* W;

(b) $(T - cI)\alpha$ *está em* W, *para algum valor característico* c *do operador* T.

Demonstração. (a) e (b) estão dizendo que o T-condutor de α em W é um polinômio linear. Seja β um vetor arbitrário em V que não esteja em W. Seja g o T-condutor de β em W. Então g divide p, o polinômio minimal de T. Como β não está em W, o polinômio g é não constante. Portanto,

$$g = (x - c_1)^{e_1} \dots (x - c_k)^{e_k}$$

onde pelo menos um dos inteiros e_i é positivo. Escolhamos j tal que $e_j > 0$. Então $(x - c_j)$ divide g:

$$g = (x - c_j)h.$$

Pela definição de g, o vetor $\alpha = h(T)\beta$ não pode estar em W. Mas

$$\begin{aligned}(T - c_jI)\alpha &= (T - c_jI)h(T)\beta \\ &= g(T)\beta\end{aligned}$$

está em W.

Teorema 5. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F *e seja* T *um operador linear sobre* V. *Então* T *é triangulável se, e somente se, o polinômio minimal de* T *é um produto de polinômios lineares sobre* F.

Demonstração. Suponhamos que o polinômio minimal se decomponha em:

$$p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}.$$

Aplicando repetidamente o lema acima, chegaremos a uma base ordenada $\mathscr{B} = \{\alpha_1, \dots, \alpha_n\}$, em relação à qual a matriz que representa T é triangular superior

$$(6\text{-}11) \qquad [T]_{\mathscr{B}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

Ora, (6-11) simplesmente diz que

$$(6\text{-}12) \qquad T\alpha_j = a_{1j}\alpha_1 + \dots + a_{jj}\alpha_j, \qquad 1 \leq j \leq n$$

isto é, $T\alpha_j$ está no subespaço gerado por $\alpha_1, \dots, \alpha_j$. Para determinar $\alpha_1, \dots, \alpha_n$, começamos aplicando o lema ao subespaço $W = \{0\}$, para obter o vetor α_1. Em seguida, aplicamos o lema a W_1, o espaço gerado por α_1 e obtemos α_2. Em seguida, aplicamos o lema a W_2, o espaço gerado por α_1 e α_2. Continuamos dessa maneira. Um ponto merece ser comentado. Após determinarmos $\alpha_1, \dots, \alpha_i$, são as relações de tipo triangular (6-12) para $j = 1, \dots, i$, que garantem que o subespaço gerado por $\alpha_1, \dots, \alpha_i$ é invariante sob T.

Se T é triangulável, é evidente que o polinômio característico de T tenha a forma

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}, \qquad c_i \text{ em } F.$$

Basta olhar para a matriz triangular (6-11). Os elementos diagonais $a_{11}, \dots, a_{nn}$ são os valores característicos, com c_i repetido d_i vezes. Mas se f pode ser fatorado dessa maneira, o mesmo acontece com o polinômio minimal p, pois p divide f.

Corolário. *Seja* F *um corpo algebricamente fechado, por exemplo, o corpo dos números complexos. Toda* n × n *matriz sobre* F *é semelhante sobre* F *a uma matriz triangular.*

Teorema 6. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F *e seja* T *um operador linear sobre* V. *Então* T *é diagonalizável se, e somente se, o polinômio minimal de* T *tem a forma*

$$p = (x - c_1) \dots (x - c_k)$$

com $c_1, \dots, c_k$ *elementos distintos de* F.

Demonstração. Observamos, anteriormente, que se T é diagonalizável, seu polinômio minimal é um produto de fatores lineares distintos (ver a discussão que precede o Exemplo 4). Para demonstrar a recíproca, seja W o subespaço gerado por todos os vetores característicos de T e suponhamos que $W \neq V$. Pelo lema usado na demonstração do Teorema 5, existe um vetor α, não em W, e um valor característico c_j de T, tal que o vetor

$$\beta = (T - c_j I)\alpha$$

esteja em W. Como β está em W,

$$\beta = \beta_1 + \dots + \beta_k$$

onde $T\beta_i = c_i\beta_i$, $1 \leq i \leq k$, e portanto, o vetor

$$h(T)\beta = h(c_1)\beta_1 + \dots + h(c_k)\beta_k$$

está em W, para todo polinômio h.

Ora, $p = (x - c_j)q$ para algum polinômio q. Também

$$q - q(c_j) = (x - c_j)h.$$

Temos

$$q(T)\alpha - q(c_j)\alpha = h(T)(T - c_j I)\alpha = h(T)\beta.$$

Mas $h(T)\beta$ está em W e, como

$$0 = p(T)\alpha = (T - c_j I)q(T)\alpha$$

o vetor $q(T)\alpha$ está em W. Portanto, $q(c_j)\alpha$ está em W. Como α não está em W, temos $q(c_j) = 0$. Isto contradiz o fato de que p possui raízes distintas.

No fim da Seção 6.7, daremos uma outra demonstração do Teorema 6. Além de ser um resultado elegante, o Teorema 6 é útil sob o aspecto computacional. Seja T um operador linear representado, em relação a alguma base ordenada, pela matriz A. Queremos saber se T é diagonalizável. Calculemos o polinômio característico f. Se pudermos decompor f:

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$$

teremos dois métodos diferentes para determinar se T é ou não diagonalizável. Um método será verificar se é possível (para cada i) determinar d_i vetores característicos independentes, associados ao valor característico c_i. O outro método será verificar se $(T - c_1 I) \dots (T - c_k I)$ é ou não o operador nulo.

O Teorema 5 oferece uma demonstração diferente do teorema de Cayley-Hamilton. Esse teorema é fácil para uma matriz triangular. Logo, por meio do Teorema 5, obteremos o resultado para uma matriz arbitrária sobre um corpo algebricamente fechado. Qualquer corpo é um subcorpo de um corpo algebricamente fechado. Se conhecermos este resultado, obteremos uma demonstração do teorema de Cayley-Hamilton para matrizes sobre um corpo arbitrário. Se aceitarmos em nossa discussão o Teorema Fundamental da Álgebra, então o Teorema 5 proporciona uma demonstração do teorema de Cayley-Hamilton para matrizes complexas e esta demonstração é independente da que foi dada anteriormente.

Exercícios

1. Seja T o operador linear sobre R^2, cuja matriz em relação à base ordenada canônica é

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}.$$

(a) Demonstrar que os únicos subespaços de R^2, que são invariantes sob T, são R^2 e o subespaço nulo.

(b) Se U é o operador linear sobre C^2, cuja matriz em relação à base ordenada canônica é A, mostrar que U possui subespaços invariantes unidimensionais.

2. Seja W um subespaço invariante sob T. Demonstrar, sem fazer uso de matrizes, que o polinômio minimal do operador T_W, restrição de T a W, divide o polinômio minimal de T.

3. Seja c um valor característico de T e seja W o espaço dos vetores característicos associado ao valor característico c. Qual é o operador T_W?

4. Seja

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 2 & -2 & 2 \\ 2 & -3 & 2 \end{bmatrix}.$$

A é semelhante sobre o corpo dos números reais a uma matriz triangular? Em caso afirmativo, determinar uma tal matriz triangular.

5. Toda matriz A, tal que $A^2 = A$, é semelhante a uma matriz diagonal.

6. Seja T um operador linear diagonalizável sobre o espaço vetorial n-dimensional V, e seja W um subespaço invariante sob T. Demonstrar que o operador T_W é diagonalizável.

7. Seja T um operador linear sobre um espaço vetorial de dimensão finita sobre o corpo dos números complexos. Demonstrar que T é diagonalizável se, e somente se, T é anulado por algum polinômio sobre C que possua raízes distintas.

8. Seja T um operador linear sobre V. Se todo subespaço de V for invariante sob T, então T será um múltiplo escalar do operador idêntico.

9. Seja T o operador integral indefinida

$$(Tf)(x) = \int_0^x f(t)\, dt$$

sobre o espaço das funções contínuas sobre o intervalo $[0, 1]$. O espaço das funções polinomiais é invariante sob T? O espaço das funções diferenciáveis? O espaço das funções que se anulam em $x = \frac{1}{2}$?

10. Seja A uma 3×3 matriz com elementos reais. Demonstrar que se A não for semelhante sobre R a uma matriz triangular, então A será semelhante sobre C a uma matriz diagonal.

11. Verdadeiro ou falso? Se a matriz triangular A for semelhante a uma matriz diagonal, então A já é diagonal.

12. Seja T um operador linear sobre um espaço vetorial de dimensão finita sobre um corpo F algebricamente fechado. Seja f um polinômio sobre F. Demonstrar que c é um valor característico de $f(T)$ se, e somente se, $c = f(t)$, onde t é um valor característico de T.

13. Seja V o espaço das $n \times n$ matrizes sobre F. Seja A uma $n \times n$ matriz fixa sobre F. Sejam T e U operadores lineares sobre V definidos por

$$T(B) = AB$$
$$U(B) = AB - BA.$$

(a) Verdadeiro ou falso? Se A for diagonalizável (sobre F), então T será diagonalizável.

(b) Verdadeiro ou falso? Se A for diagonalizável, então U será diagonalizável.

6.5 Triangulação Simultânea; Diagonalização Simultânea

Seja V um espaço de dimensão finita e seja $\mathcal{F}$ uma família de operadores lineares sobre V. Queremos saber quando é possível triangular ou diagonalizar simultaneamente os operadores em $\mathcal{F}$, isto é, determinar uma base $\mathcal{B}$ tal que todas as matrizes $[T]_{\mathcal{B}}$, T em $\mathcal{F}$, sejam triangulares (ou diagonais). No caso da diagonalização é necessário que $\mathcal{F}$ seja uma família comutativa de operadores: $UT = TU$ para todo T e u em $\mathcal{F}$. Isso decorre do fato de que todas as matrizes diagonais comutam. É claro que também é necessário que cada operador em $\mathcal{F}$ seja um operador diagonalizável. A fim de triangularmos simultaneamente, cada operador em $\mathcal{F}$ terá que ser triangulável. Não é necessário que $\mathcal{F}$ seja uma família comutativa; no entanto, esta condição é suficiente para triangulação simultânea (se cada T for individualmente triangulável). Esses resultados decorrem de pequenas variações das demonstrações dos Teoremas 5 e 6.

O subespaço W é **invariante sob** (a família de operadores) $\mathcal{F}$ se W é invariante sob cada operador em $\mathcal{F}$.

Lema. *Seja $\mathcal{F}$ uma família comutativa de operadores lineares trianguláveis sobre* V. *Seja* W *um subespaço próprio de* V, *invariante sob $\mathcal{F}$. Então existe um vetor α em* V *tal que*

(a) *α não está em* W;

(b) *para cada* T *em $\mathcal{F}$, o vetor* Tα *está no subespaço gerado por α e* W.

Demonstração. Não há perda de generalidade se admitirmos que $\mathcal{F}$ contenha apenas um número finito de operadores, por causa da seguinte observação: Seja $\{T_1, \ldots, T_r\}$ um subconjunto linearmente independente maximal de $\mathcal{F}$, isto é, uma base do subespaço gerado por $\mathcal{F}$. Se α é um vetor tal que (b) é verdadeiro para cada T_i, então (b) será verdadeiro para todo operador que seja uma combinação linear de $T_1, \ldots, T_r$.

Pelo lema que precede o Teorema 5 (o lema atual para um só operador), podemos determinar um vetor β_1 (não em W) e um escalar c_1, tal que $(T_1 - c_1 I)\beta_1$ esteja em W. Seja V_1 a coleção de todos os vetores β em V tais que $(T_1 - c_1 I)\beta$ esteja em W. Então

V_1 é um subespaço de V propriamente maior que W. Além do mais, V_1 é invariante sob $\mathscr{F}$ pela seguinte razão: Se T comutar com T_1, então

$$(T_1 - c_1 T)(T\beta) = T(T_1 - c_1 I)\beta.$$

Se β está em V_1, então $(T_1 - c_1 I)\beta$ estará em W. Como W é invariante sob cada T em $\mathscr{F}$, teremos $T(T_1 - c_1 I)\beta$ em W, isto é, $T\beta$ em V_1 para todo β em V_1 e todo T em $\mathscr{F}$.

Ora, W é um subespaço próprio de V_1. Seja U_2 o operador linear sobre V_1 obtido pela restrição de T_2 ao subespaço V_1. O polinômio minimal de U_2 divide o polinômio minimal de T_2. Portanto, podemos aplicar o lema que precede o Teorema 5 a esse operador e ao subespaço invariante W. Obteremos um vetor β_2 em V_1 (não em W) e um escalar c_2 tais que $(T_2 - c_2 I)\beta_2$ esteja em W. Observemos que

(a) β_2 não está em W;

(b) $(T_1 - c_1 I)\beta_2$ está em W;

(c) $(T_2 - c_2 I)\beta_2$ está em W.

Seja V_2 o conjunto de todos os vetores β em V_1 tais que $(T_2 - c_2 I)\beta$ esteja em W. Então V_2 é invariante sob $\mathscr{F}$. Apliquemos o lema que precede o Teorema 5 a U_3, a restrição de T_3 a V_2. Continuando dessa maneira, chegaremos a um vetor $\alpha = \beta_r$ (não em W) tal que $(T_j - c_j I)\alpha$ esteja em W, $j = 1, \ldots, r$.

Teorema 7. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F. *Seja* $\mathscr{F}$ *uma família comutativa de operadores lineares trianguláveis sobre* V. *Existe uma base ordenada de* V *tal que todo operador em* $\mathscr{F}$ *seja representado, em relação àquela base, por uma matriz triangular.*

Demonstração. Dado o lema que acabamos de provar, a demonstração deste teorema é igual à demonstração do Teorema 5 colocando T no lugar de $\mathscr{F}$.

Corolário. *Seja* $\mathscr{F}$ *uma família comutativa de* n × n *matrizes sobre um corpo* F *algebricamente fechado. Existe uma* n × n *matriz* P, *não-singular, com elementos em* F, *tal que* $P^{-1}AP$ *seja triangular-superior para toda matriz* A *em* $\mathscr{F}$.

Teorema 8. *Seja* $\mathscr{F}$ *uma família comutativa de operadores lineares diagonalizáveis sobre o espaço vetorial* V *de dimensão finita.*

Então existe uma base ordenada de V *tal que todo operador em* $\mathscr{F}$ *seja representado, em relação àquela base, por uma matriz diagonal.*

Demonstração. Poderíamos demonstrar esse teorema adaptando o lema que precede o Teorema 7 ao caso diagonalizável, exatamente como adaptamos o lema que precede o Teorema 5 ao caso diagonalizável para demonstrar o Teorema 6. No entanto, nesse ponto, é mais fácil procedermos por indução sobre a dimensão de V.

Se dim $V = 1$, não há nada a demonstrar. Admitamos o teorema para espaços vetoriais de dimensão menor que n, e seja V um espaço n-dimensional. Escolhamos um T arbitrário em $\mathscr{F}$ que não seja múltiplo escalar do operador idêntico. Sejam $c_1, \ldots, c_k$ os valores característicos distintos de T, e (para cada i) seja W_i o núcleo de $T - c_iI$. Fixemos um índice i. Então W_i será invariante sob cada operador que comute com T. Seja $\mathscr{F}_i$ a família de operadores lineares sobre W_i, obtida pela restrição dos operadores em $\mathscr{F}$ ao subespaço (invariante) W_i. Cada operador em $\mathscr{F}_i$ é diagonalizável porque o seu polinômio minimal divide o polinômio minimal do operador correspondente em $\mathscr{F}$. Como dim $W_i <$ dim V, os operadores em $\mathscr{F}_i$ podem ser diagonalizados simultaneamente. Em outras palavras, W_i possui uma base $\mathscr{B}_i$ que consiste de vetores que são, simultaneamente, vetores característicos de todos os operadores em $\mathscr{F}_i$.

Como T é diagonalizável, o lema que precede o Teorema 2, nos diz que $\mathscr{B} = (\mathscr{B}_1, \ldots, \mathscr{B}_k)$ é uma base de V. Esta é a base que procuramos.

Exercícios

1. Determinar uma matriz real inversível P tal que $P^{-1}AP$ e $P^{-1}BP$ sejam ambas diagonalizáveis, onde A e B são as matrizes reais

(a) $$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -8 \\ 0 & -1 \end{bmatrix}$$

(b) $$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}.$$

2. Seja $\mathscr{F}$ uma família comutativa de 3×3 matrizes complexas. Quantas matrizes linearmente independentes existem em $\mathscr{F}$? E no caso $n \times n$?

3. Seja T um operador linear sobre um espaço n-dimensional, e suponhamos que T possua n valores característicos distintos. Demonstrar que qualquer operador linear que comuta com T é um polinômio em T.

4. Sejam A, B, C e D $n \times n$ matrizes complexas que comutam. Seja E a $2n \times 2n$ matriz

$$E = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Demonstrar que $\det E = \det(AD - BC)$.

5. Seja F um corpo, n um inteiro positivo e seja V o espaço das $n \times n$ matrizes sobre F. Se A é uma $n \times n$ matriz fixa sobre F, seja T_A o operador linear sobre V definido por $T_A(B) = AB - BA$. Consideremos a família dos operadores lineares T_A obtida fazendo A percorrer o conjunto das matrizes diagonais. Demonstrar que os operadores desta família são simultaneamente diagonalizáveis.

6.6 Decomposição em Somas Diretas

Ao continuarmos nossa análise de um só operador linear, formularemos nossas idéias de uma maneira ligeiramente mais sofisticada — menos em termos de matrizes e mais em termos de subespaços. Quando iniciamos esse capítulo, descrevemos nossos objetivos dessa maneira: determinar uma base ordenada em relação à qual a matriz de T assumisse uma forma especialmente simples. Agora descreveremos nosso objetivo assim: Decompor o espaço subjacente V em uma soma de subespaços invariantes sob T, tais que as restrições dos operadores a esses subespaços sejam simples.

Definição. *Sejam* $W_1, \ldots, W_k$ *subespaços do espaço vetorial* V. *Diremos que* $W_1, \ldots, W_k$ *são* **independentes** *se*

$$\alpha_1 + \ldots + \alpha_k = 0, \quad \alpha_i \text{ em } W_i$$

implica que cada α_i *é nulo.*

Para $k = 2$, o significado da independência é "interseção igual a $\{0\}$", isto é, W_1 e W_2 são independentes se, e somente se, $W_1 \cap W_2 = \{0\}$. Se $k > 2$, a independência de $W_1, \ldots, W_k$ diz muito mais do que $W_1 \cap \ldots \cap W_k = \{0\}$. Ela diz que a interseção de cada W_j com a soma dos outros subespaços W_i contém somente o vetor nulo.

O significado da independência é o seguinte: Seja $W = W_1 + \ldots + W_k$ o subespaço gerado por $W_1, \ldots, W_k$. Cada vetor α em W pode ser expresso como uma soma

$$\alpha = \alpha_1 + \ldots + \alpha_k \qquad \alpha_i \text{ em } W_i.$$

Se $W_1, \ldots, W_k$ forem independentes, então a expressão para α será única; de fato, se

$$\alpha = \beta_1 + \ldots + \beta_k \qquad \beta_i \text{ em } W_i$$

então $0 = (\alpha_1 - \beta_1) + \ldots + (\alpha_k - \beta_k)$, logo, $\alpha_i - \beta_i = 0$, $i = 1, \ldots, k$. Assim, quando $W_1, \ldots, W_k$ são independentes, podemos operar com os vetores de W como k-uplas $(\alpha_1, \ldots, \alpha_k)$, α_i em W_i, da mesma maneira como operamos com vetores em R^k como k-uplas de números.

Lema. *Seja* V *um espaço vetorial de dimensão finita. Sejam* $\mathrm{W}_1, \ldots, \mathrm{W}_k$ *subespaços de* V *e seja* $\mathrm{W} = \mathrm{W}_1 + \ldots + \mathrm{W}_k$. *As seguintes condições são equivalentes:*

(a) $\mathrm{W}_1, \ldots, \mathrm{W}_k$ *são independentes.*

(b) *Para cada* j, $2 \leq \mathrm{j} \leq \mathrm{k}$, *temos*

$$\mathrm{W}_j \cap (\mathrm{W}_1 + \ldots + \mathrm{W}_{j-1}) = \{0\}.$$

(c) *Se* $\mathscr{B}_i$ *é uma base ordenada de* W_i, $1 \leq \mathrm{i} \leq \mathrm{k}$, *então a sequência* $\mathscr{B} = (\mathscr{B}_1, \ldots, \mathscr{B}_k)$ *é uma base ordenada de* W.

Demonstração. Suponhamos (a). Seja α um vetor na interseção $W_j \cap (W_1 + \ldots + W_{j-1})$. Então existem vetores $\alpha_1, \ldots, \alpha_{j-1}$, com α_i em W_i, tais que $\alpha = \alpha_1 + \ldots + \alpha_{j-1}$. Como

$$\alpha_1 + \ldots + \alpha_{j-1} + (-\alpha) + 0 + \ldots + 0 = 0$$

e como $W_1, \ldots, W_k$ são independentes, necessariamente $\alpha_1 = \alpha_2 = \ldots = \alpha_{j-1} = \alpha = 0$.

Observemos agora que (b) implica (a). Suponhamos

$$0 = \alpha_1 + \ldots + \alpha_k \qquad \alpha_i \text{ em } W_i.$$

Seja j o maior inteiro i tal que $\alpha_i \neq 0$. Então

$$0 = \alpha_1 + \ldots + \alpha_j \qquad \alpha_j \neq 0.$$

Logo, $\alpha_j = -\alpha_1 - \ldots - \alpha_{j-1}$ é um vetor não nulo em $W_j \cap (W_1 + \ldots + W_{j-1})$.

Agora que sabemos que (a) e (b) são equivalentes, vejamos por que (a) é equivalente a (c). Suponhamos (a). Seja $\mathscr{B}_i$ uma base de W_i, $1 \leq i \leq k$ e seja $\mathscr{B} = (\mathscr{B}_1, \ldots, \mathscr{B}_k)$. Qualquer relação linear entre os vetores de $\mathscr{B}$ terá a forma

$$\beta_1 + \ldots + \beta_k = 0$$

onde β_i é alguma combinação linear dos vetores de $\mathscr{B}_i$. Como $W_1, \ldots, W_k$ são independentes, cada β_i é 0. Como cada $\mathscr{B}$ é independente, a relação existente entre os vetores de $\mathscr{B}$ é a relação trivial.

Deixamos a demonstração de que (c) implica (a) como exercício (Exercício 2).

Se uma (e portanto todas) das três condições do último lema é válida, diremos que a soma $W = W_1 + \ldots + W_k$ é **direta** ou que W é a **soma direta** de $W_1, \ldots, W_k$ e escreveremos

$$W = W_1 \oplus \ldots \oplus W_k.$$

Na literatura, o leitor poderá encontrar essa soma direta sob os nomes de soma independente ou soma interna de $W_1, \ldots, W_k$.

Exemplo 11. Seja V um espaço vetorial de dimensão finita sobre o corpo F e seja $\{\alpha_1, \ldots, \alpha_n\}$ uma base arbitrária de V. Se W_i é o subespaço unidimensional gerado por α_i, então $V = W_1 \oplus \ldots \oplus W_n$.

Exemplo 12. Seja n um inteiro positivo, F um subcorpo do corpo dos números complexos e seja V o espaço das $n \times n$ matrizes sobre F. Seja W_1 o subespaço das matrizes **simétricas**, isto é, matrizes A tais que $A^t = A$. Seja W_2 o subespaço das matrizes **anti-simétricas**, isto é, matrizes tais que $A^t = -A$. Então $V = W_1 \otimes W_2$. Se A é uma matriz arbitrária em V, a única expressão para A como uma soma de duas matrizes, uma em W_1 e a outra em W_2, é

$$\begin{aligned} A &= A_1 + A_2 \\ A_1 &= \tfrac{1}{2}(A + A^t) \\ A_2 &= \tfrac{1}{2}(A - A^t). \end{aligned}$$

Exemplo 13. Seja T um operador linear arbitrário sobre um espaço vetorial V de dimensão finita. Sejam $c_1, \ldots, c_k$ valores característicos distintos de T e seja W_i o espaço dos vetores característicos associado ao valor característico c_i. Então, $W_1, \ldots, W_k$ são independentes. Ver o lema que precede o Teorema 2. Em particular, se T é diagonalizável, então $V = W_1 \oplus \ldots \oplus W_k$.

Definição. *Se* V *é um espaço vetorial, uma* **projeção** *de* V *é um operador linear* E *sobre* V *tal que* $E^2 = E$.

Suponhamos que E seja uma projeção. Sejam R a imagem de E e N o núcleo de E.

1. O vetor β está na imagem R se, e somente se, $E\beta = \beta$. Se $\beta = E\alpha$, então $E\beta = E^2\alpha = E\alpha = \beta$. Reciprocamente, se $\beta = E\beta$, então (é claro) β está na imagem de E.

2. $V = R \oplus N$.

3. A única expressão de α como uma soma de vetores em R e N é $\alpha = E\alpha + (\alpha - E\alpha)$.

De (1), (2) e (3) é fácil ver o seguinte: Se R e N são subespaços de V tais que $V = R \oplus N$, então existe um e um só operador projeção E cuja imagem seja R e cujo núcleo seja N. Esse operador é chamado **projeção sobre R segundo N**.

Qualquer projeção E é (trivialmente) diagonalizável. Se $\{\alpha_1, \ldots, \alpha_r\}$ é uma base de R e $\{\alpha_{r+1}, \ldots, \alpha_n\}$ uma base de N, então a base $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ diagonaliza E:

$$[E]_{\mathscr{B}} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

onde I é a $r \times r$ matriz unidade. Isso deveria esclarecer parte da terminologia ligada a projeções. O leitor deverá examinar vários casos no plano R^2 (ou no espaço tri-dimensional R^3) para se convencer de que a projeção sobre R segundo N leva cada vetor para R, projetando-o paralelamente a N.

Projeções podem ser usadas para descrever decomposições do espaço V em somas diretas. De fato, suponhamos $V = W_1 \oplus \ldots \oplus W_k$. Para cada j definiremos um operador E_j sobre V. Seja α em V, digamos $\alpha = \alpha_1 + \ldots + \alpha_k$ com α_i em W_i. Definamos $E_j\alpha = \alpha_j$. Então, E_j é uma regra bem definida. É fácil ver que E_j é linear, que a imagem de E_j é W_j e que $E_j^2 = E_j$. O núcleo de E_j é o subespaço

$$(W_1 + \ldots + W_{j-1} + W_{j+1} + \ldots + W_k)$$

pois a afirmação de que $E_j\alpha = 0$ significa simplesmente que $\alpha_j = 0$, isto é, que α é na realidade uma soma de vetores dos espaços W_i com $i \neq j$. Em termos das projeções E_j temos

(6-13) $$\alpha = E_1\alpha + \ldots + E_k\alpha$$

para cada α em V. O que (6-13) diz é que

$$I = E_1 + \dots + E_k.$$

Notemos também que se $i \neq j$ então $E_i E_j = 0$, pois a imagem de E_j é o subespaço W_j que está contido no núcleo de E_i. Resumiremos agora nossas conclusões, enunciaremos e demonstraremos uma recíproca.

Teorema 9. *Se* $V = W_1 \oplus \dots \oplus W_k$, *então existem* k *operadores* $E_1, \dots, E_k$ *sobre* V *tais que*

(i) *cada* E_i *é uma projeção* ($E_i^2 = E_i$);

(ii) $E_i E_j = 0$, *se* $i \neq j$;

(iii) $I = E_1 + \dots + E_k$;

(iv) *a imagem de* E_i *é* W_i.

Reciprocamente, se $E_1, \dots, E_k$ *são* k *operadores lineares sobre* V *que satisfazem as condições* (i), (ii) *e* (iii) *e se indicamos por* W_i *a imagem de* E_i, *então* $V = W_1 \oplus \dots \oplus W_k$.

Demonstração. Precisamos demonstrar apenas a afirmação recíproca. Suponhamos que $E_1, \dots, E_k$ sejam operadores lineares sobre V que satisfaçam as três primeiras condições e seja W_i a imagem de E_i. Então, certamente

$$V = W_1 + \dots + W_k;$$

pois, pela condição (iii) temos

$$\alpha = E_1\alpha + \dots + E_k\alpha$$

para cada α em V e $E_i\alpha$ está em W_i. Esta expressão para α é única, porque se

$$\alpha = \alpha_1 + \dots + \alpha_k$$

com α_i em W_i, digamos $\alpha_i = E_i\beta_i$, então, usando (i) e (ii), temos

$$\begin{aligned} E_j\alpha &= \sum_{i=1}^{k} E_j\alpha_i \\ &= \sum_{i=1}^{k} E_j E_i \beta_i \end{aligned}$$

$$= E_j^2 \beta_j$$
$$= E_j \beta_j$$
$$= \alpha_j.$$

Isto mostra que V é a soma direta dos W_i.

Exercícios

1. Seja V um espaço vetorial de dimensão finita e seja W_1 um subespaço arbitrário de V. Demonstrar que existe um subespaço W_2 de V tal que $V = W_1 \oplus W_2$.

2. Seja V um espaço vetorial de dimensão finita e sejam $W_1, \ldots, W_k$ subespaços de V tais que

$$V = W_1 + \ldots + W_k \text{ e } \dim V = \dim W_1 + \ldots + \dim W_k.$$

Demonstrar que $V = W_1 \oplus \ldots \oplus W_k$.

3. Determinar uma projeção E que projeta R^2 sobre o subespaço gerado por $(1, -1)$ segundo o subespaço gerado por $(1, 2)$.

4. Se E_1 e E_2 são projeções sobre subespaços independentes então $E_1 + E_2$ é uma projeção. Verdadeiro ou falso?

5. Se E é uma projeção e f um polinômio, então $f(E) = aI + bE$. O que são a e b em termos dos coeficientes de f?

6. Verdadeiro ou falso? Se os únicos valores característicos de um operador diagonalizável são 0 e 1, então esse operador é uma projeção.

7. Demonstrar que se E é a projeção sobre R segundo N, então $(I - E)$ é a projeção sobre N segundo R.

8. Sejam $E_1, \ldots, E_k$ operadores lineares sobre o espaço V tais que $E_1 + \ldots + E_k = I$.

   (a) Demonstrar que se $E_i E_j = 0$ para $i \neq j$, então $E_i^2 = E_i$ para todo i.

   (b) No caso $k = 2$, demonstrar a recíproca de (a). Isto é, se $E_1 + E_2 = I$ e $E_1^2 = E_1$, $E_2^2 = E_2$, então $E_1 E_2 = 0$.

9. Seja V um espaço vetorial real e E um operador linear idempotente sobre V, isto é, uma projeção. Demonstrar que $(I + E)$ é inversível. Determinar $(I + E)^{-1}$.

10. Seja F um subcorpo do corpo dos números complexos (ou, um corpo de característica zero). Seja V um espaço vetorial de dimensao finita sobre F. Suponhamos que $E_1, \ldots, E_k$ sejam projeções sobre V e que $E_1 + \ldots + E_k = I$. Demonstrar que $E_i E_j = 0$ para $i \neq j$. (*Sugestão*: Usar a função traço e descobrir o que é o traço de uma projeção.)

11. Seja V um espáço vetorial, sejam $W_1, \ldots, W_k$ subespaços de V e seja

$$V_j = W_1 + \ldots + W_{j-1} + W_{j+1} + \ldots + W_k$$

Suponhamos que $V = W_1 \oplus \ldots \oplus W_k$. Demonstrar que o espaço dual V^* admite a decomposição em soma direta $V^* = V_1^0 \oplus \ldots \oplus V_k^0$.

6.7 Somas Diretas Invariantes

Estamos essencialmente interessados em decomposições em somas diretas $V = W_1 \oplus \ldots \oplus W_k$, onde cada um dos subespaços W_i seja invariante sob algum operador linear dado T. Dada uma tal decomposição de V, T induz um operador linear T_i, sobre cada W_i, através da restrição. A ação de T é então a que segue. Se α é um vetor em V, existem vetores bem determinados $\alpha_1, \ldots, \alpha_k$ com α_i em W_i, tais que

$$\alpha = \alpha_1 + \ldots + \alpha_k$$

é então

$$T\alpha = T_1\alpha_1 + \ldots + T_k\alpha_k.$$

Descreveremos esta situação dizendo que T é a **soma direta** dos operadores $T_1, \ldots, T_k$. Deve-se lembrar, ao usarmos esta terminologia, que os T_i não são operadores lineares sobre o espaço V mas sim sobre os diversos subespaços W_i. O fato de que $V = W_1 \oplus \ldots \oplus W_k$ nos permite associar a cada α em V uma única k-upla $(\alpha_1, \ldots, \alpha_k)$ de vetores α_i em W_i (sendo $\alpha = \alpha_1 + \ldots + \alpha_k$) de uma maneira tal que possamos efetuar as operações lineares em V trabalhando em cada subespaço W_i. O fato de que cada W_i é invariante sob T nos permite considerar a ação de T como a ação independente dos operadores T_i sobre os subespaços W_i. Nosso propósito é estudar T determinando decomposições em somas diretas invariantes nas quais os T_i sejam operadores de natureza elementar.

Antes de considerarmos um exemplo, observemos o análogo desta situação para matrizes. Suponhamos que tomemos uma base ordenada $\mathscr{B}_i$ para cada W_i e seja $\mathscr{B}$ a base ordenada de V formada pela reunião das $\mathscr{B}_i$, ordenada como $\mathscr{B}_1, \ldots, \mathscr{B}_k$, de modo que $\mathscr{B}$ seja uma base de V. De nossa discussão acima, quanto ao análogo para matrizes para um único subespaço invariante, é fácil ver que se $A = [T]_{\mathscr{B}}$ e $A_i = [T_i]_{\mathscr{B}_i}$, então A é da forma em blocos

(6-14)
$$A = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & A_k \end{bmatrix}.$$

Em (6-14), A_i é uma $d_i \times d_i$ matriz ($d_i = \dim W_i$) e os símbolos 0 são blocos retangulares de vários tamanhos constituídos de escalares nulos. Parece ser também apropriado descrever (6-14) dizendo que A é a **soma direta** das matrizes $A_1, \dots, A_k$.

Na maioria das vezes descreveremos o subespaço W_i por meio das projeções associadas E_i (Teorema 9). Portanto, teremos que formular a invariância dos subespaços W_i em termos dos E_i.

Teorema 10. *Seja* T *um operador linear sobre o espaço* V *e sejam* $W_1, \dots, W_k$ *e* $E_1, \dots, E_k$ *como no Teorema* 9. *Então, uma condição necessária e suficiente para que cada subespaço* W_i *seja invariante sob* T *é que* T *comute com cada uma das projeções* E_i, *isto* é,

$$TE_i = E_iT, \; i = 1, \dots, k.$$

Demonstração. Suponhamos que T comute com cada E_i. Seja α em W_j. Então $E_j\alpha = \alpha$ e

$$\begin{aligned} T\alpha &= T(E_j\alpha) \\ &= E_j(T\alpha) \end{aligned}$$

o que mostra que $T\alpha$ está na imagem de E_j, isto é, que W_j é invariante sob T.

Suponhamos agora que cada W_i seja invariante sob T. Mostraremos que $TE_j = E_jT$. Seja α um vetor arbitrário em V. Então

$$\begin{aligned} \alpha &= E_1\alpha + \dots + E_k\alpha \\ T\alpha &= TE_1\alpha + \dots + TE_k\alpha. \end{aligned}$$

Como $E_i\alpha$ está em W_i, que é invariante sob T, devemos ter $T(E_i\alpha) = E_i\beta_i$ para algum vetor β_i. Então

$$E_jTE_i\alpha = E_jE_i\beta_i$$

$$= \begin{cases} 0, & \text{se } i \neq j \\ E_j\beta_j, & \text{se } i = j. \end{cases}$$

Assim,

$$\begin{aligned} E_jT\alpha &= E_jTE_1\alpha + \ldots + E_jTE_k\alpha \\ &= E_j\beta_j \\ &= TE_j\alpha. \end{aligned}$$

Isto vale para todo α em V, portanto $E_jT = TE_j$.

Descreveremos agora um operador diagonalizável T utilizando decomposições em somas diretas invariantes (projeções que comutam com T). Isso será de grande ajuda para compreendermos mais tarde alguns teoremas de decomposição mais profundos. O leitor talvez ache complicada a descrição que iremos dar, em comparação com a formulação por meio de matrizes ou com a afirmação de que os vetores característicos de T geram o espaço subjacente. Mas, ele deverá ter em mente que este é nosso primeiro contato com um método muito eficiente, graças ao qual vários problemas ligados a subespaços, bases, matrizes e outros semelhantes, poderão ser reduzidos a cálculos algébricos com operadores lineares. Com um pouco de experiência, a eficiência e elegância desse modo de pensar tornar-se-á aparente.

Teorema 11. *Seja* T *um operador linear sobre um espaço* V *de dimensão finita. Se* T *é diagonalizável e* $c_1, \ldots, c_k$ *são os valores característicos distintos de* T, *então existem operadores lineares* $E_1, \ldots, E_k$ *sobre* V *tais que*

(i) $T = c_1E_1 + \ldots + c_kE_k$;
(ii) $I = E_1 + \ldots + E_k$;
(iii) $E_iE_j = 0$, $i \neq j$;
(iv) $E_i^2 = E_i$; (E_i *é uma projeção*);
(v) *a imagem de* E_i *é o espaço característico de* T *associado a* c_i.

Reciprocamente, se existirem k *escalares distintos* $c_1, \ldots, c_k$ *e* k *operadores lineares não-nulos* $E_1, \ldots, E_k$ *satisfazendo as condições* (i), (ii) e (iii), *então* T *é diagonalizável,* $c_1, \ldots, c_k$ *são os valores característicos distintos de* T *e as condições* (iv) *e* (v) *também são satisfeitas.*

Demonstração. Suponhamos que T seja diagonalizável, com valores característicos distintos $c_1, \ldots, c_k$. Seja W_i o espaço dos vetores característicos associado ao valor característico c_i. Como vimos

$$V = W_1 \oplus \ldots \oplus W_k.$$

Sejam $E_1, \ldots, E_k$ as projeções associadas a esta decomposição, como no Teorema 9. Então (ii), (iii), (iv) e (v) são satisfeitas. Para verificar (i) procedamos da seguinte maneira: para cada α em V,

$$\alpha = E_1\alpha + \ldots + E_k\alpha$$

logo

$$\begin{aligned} T\alpha &= TE_1\alpha + \ldots + TE_k\alpha \\ &= c_1E_1\alpha + \ldots + c_kE_k\alpha. \end{aligned}$$

Em outras palavras $T = c_1E_1 + \ldots + c_kE_k$.

Suponhamos agora que sejam dados um operador linear T juntamente com escalares distintos c_i e operadores não-nulos E_i que satisfaçam (i), (ii) e (iii). Como $E_iE_j = 0$ para $i \neq j$, multiplicando ambos os membros de $I = E_1 + \ldots + E_k$ por E_i obtemos imediatamente $E_i^2 = E_i$. Multiplicando $T = c_1E_1 + \ldots + c_kE_k$ por E_i resulta $TE_i = c_iE_i$, o que mostra que todo vetor na imagem de E_i está no núcleo de $(T - c_iI)$. Como supusemos que $E_i \neq 0$, isto demonstra que existe um vetor não-nulo no núcleo de $(T - c_iI)$, isto é, que c_i e um valor característico de T. Além disso, os c_i são os únicos valores característicos de T; de fato, se c é um escalar arbitrário, então

$$T - cI = (c_1 - c)E_1 + \ldots + (c_k - c)E_k$$

portanto, se $(T - cI)\alpha = 0$, devemos ter $(c_i - c)E_i\alpha = 0$. Se α não é o vetor nulo então $E_i\alpha \neq 0$ para algum i, de modo que para este i temos $c_i - c = 0$.

T é certamente diagonalizável, pois mostramos que todo vetor não-nulo na imagem de E_i é um vetor característico de T e o fato de que $I = E_1 + \ldots + E_k$ mostra que esses vetores característicos geram V. Tudo o que resta a ser demonstrado é que o núcleo de $(T - c_iI)$ é exatamente a imagem de E_i. Mas isto é evidente, porque se $T\alpha = c_i\alpha$, então

$$\sum_{j=1}^{k} (c_j - c_i)E_j\alpha = 0$$

logo

$$(c_j - c_i)E_j\alpha = 0 \text{ para cada } j$$

e então

$$E_j\alpha = 0, \quad j \neq i.$$

Como $\alpha = E_1\alpha + \ldots + E_k\alpha$, e $E_j\alpha = 0$ para $j \neq i$ temos $\alpha = E_i\alpha$, demonstrando que α está na imagem de E_i.

Uma parte do Teorema 9 diz que para um operador diagonalizável T, os escalares $c_1, \ldots, c_k$ e os operadores $E_1, \ldots, E_k$ são determinados de modo único pelas condições (i), (ii), (iii), mais o fato de que os c_i são distintos e o fato de que os E_i são não-nulos. Uma das características agradáveis da decomposição $T = c_1E_1 + \ldots + c_kE_k$ é que g é um polinômio arbitrário sobre o corpo F, então

$$g(T) = g(c_1)E_1 + \ldots + g(c_k)E_k.$$

Deixemos os detalhes da demonstração a cargo do leitor. Para ver como se demonstra, basta calcular T^r para todo inteiro positivo r. Por exemplo

$$\begin{aligned} T^2 &= \sum_{i=1}^{k} c_iE_i \sum_{j=1}^{k} c_jE_j \\ &= \sum_{i=1}^{k} \sum_{j=1}^{k} c_ic_jE_iE_j \\ &= \sum_{i=1}^{k} c_i^2E_i^2 \\ &= \sum_{i=1}^{k} c_i^2E_i. \end{aligned}$$

O leitor deve comparar isto com $g(A)$ sendo A uma matriz diagonal, pois neste caso $g(A)$ é simplesmente a matriz diagonal de elementos diagonais $g(A_{11}), \ldots, g(A_{nn})$.

Gostaríamos de observar em particular o que acontece no caso dos polinômios de Lagrange correspondentes aos escalares $c_1, \ldots, c_k$:

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}.$$

Temos $p_j(c_i) = \delta_{ij}$, o que significa que

$$p_j(T) = \sum_{i=1}^{k} \delta_{ij} E_i$$
$$= E_j.$$

Assim, as projeções E_j não apenas comutam com T, mas são polinômios em T.

Cálculos como este, com polinômios em T, podem ser usados para dar uma outra demonstração do Teorema 6 que caracterizou operadores diagonalizáveis em termos de seus polinômios minimais. A demonstração é inteiramente independente da nossa demonstração anterior.

Se T é diagonalizável, $T = c_1 E_1 + \ldots + c_k E_k$, então

$$g(T) = g(c_1)E_1 + \ldots + g(c_k)E_k$$

para todo polinômio g. Logo, $g(T) = 0$ se, e somente se, $g(c_i) = 0$ para cada i. Em particular, o polinômio minimal de T é

$$p = (x - c_1) \ldots (x - c_k).$$

Suponhamos agora que T seja um operador linear sobre V com polinômio minimal

$$p = (x - c_1) \ldots (x - c_k)$$

onde $c_1, \ldots, c_k$ são elementos distintos do corpo dos escalares. Formemos os polinômios de Lagrange

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}.$$

Recordemos do Capítulo 4 que $p_j(c_i) = \delta_{ij}$ e, para todo polinômio g de grau menor ou igual a $(k-1)$, temos

$$g = g(c_1)p_1 + \ldots + g(c_k)p_k.$$

Tomando g como sendo o polinômio constante 1 e logo em seguida como o polinômio x, temos

$$\text{(6-15)} \qquad \begin{aligned} 1 &= p_1 + \ldots + p_k \\ x &= c_1 p_1 + \ldots + c_k p_k. \end{aligned}$$

(O leitor astuto terá notado que a aplicação a x pode não ser válida porque k pode ser igual a 1. Mas se $k = 1$, T é um múltiplo escalar do operador idêntico, sendo portanto diagonalizável.) Seja agora $E_j = p_j(T)$. De (6-15) temos

$$\begin{aligned} I &= E_1 + \dots + E_k \\ T &= c_1E_1 + \dots + c_kE_k. \end{aligned} \tag{6-16}$$

Observemos que se $i \neq j$ então p_ip_j é divisível pelo polinômio minimal p, pois p_ip_j contém cada $(x - c_r)$ como um fator. Assim

$$E_iE_j = 0, \quad i \neq j. \tag{6-17}$$

Precisamos notar ainda mais um fato, a saber, que $E_i \neq 0$ para todo i. Isto ocorre porque p é o polinômio minimal de T e então não podemos ter $p_i(T) = 0$, pois p_i tem grau menor que o de p. Este último comentário, junto com (6-16), (6-17) e o fato de que os c_i são distintos nos permite aplicar o Teorema e concluir que T é diagonalizável.

Exercícios

1. Seja E uma projeção sobre V e seja T um operador linear sobre V. Demonstrar que a imagem de E é invariante sob T se, e somente se, $ETE = TE$. Demonstrar que a imagem e o núcleo de E são invariantes sob T se, e somente se, $ET = TE$.

2. Seja T o operador linear sobre R^2, cuja matriz em relação à base ordenada canônica é

$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$$

Seja W_1 o subespaço de R^2 gerado pelo vetor $\varepsilon_1 = (1, 0)$.

(a) Demonstrar que W_1 é invariante sob T.

(b) Demonstrar que não existe nenhum subespaço W_2 que seja invariante sob T e que seja complementar de W_1:

$$R^2 = W_1 \oplus W_2.$$

(Comparar com o Problema 1 da Seção 6.5.)

3. Seja T um operador linear sobre o espaço vetorial V de dimensão finita, seja R a imagem de T e seja N o núcleo de T. Demonstrar que R e N são independentes se, e somente se, $V = R \oplus N$.

4. Seja T um operador linear sobre V. Suponhamos $V = W_1 \oplus \ldots \oplus W_k$, onde cada W_i é invariante sob T. Seja T_i a restrição do operador T a W_i.

   (a) Demonstrar que $\det(T) = \det(T_1) \ldots \det(T_k)$.

   (b) Demonstrar que o polinômio característico de f é o produto dos polinômios característicos de $f_1, \ldots, f_k$.

   (c) Demonstrar que o polinômio minimal de T é o mínimo múltiplo comum dos polinômios minimais de $T_1, \ldots, T_k$. (*Sugestão*: Demonstrar e em seguida usar os fatos correspondentes a respeito de somas diretas de matrizes.)

5. Seja T o operador linear, diagonalizável sobre R^3, discutido no Exemplo 3 da Seção 6.2. Usar os polinômios de Lagrange para escrever a matriz A, que representa T, na forma $A = E_1 + 2E_2$, $E_1 + E_2 = I$, $E_1E_2 = 0$.

6. Seja A a 4×4 matriz do Exemplo 6 da Seção 6.3. Determinar matrizes E_1, E_2, E_3 tais que $A = c_1E_1 + c_2E_2 + c_3E_3$, $E_1 + E_2 + E_3 = I$ e $E_iE_j = 0$, $i \neq j$.

7. Nos Exercícios 5 e 6 observar que (para cada i) o espaço dos vetores característicos associado ao valor característico c_i é gerado pelos vetores-coluna das várias matrizes E_j com $j \neq i$. Isso é uma coincidência?

8. Seja T um operador linear sobre V que comuta com todos os operadores projeção sobre V. O que se pode afirmar sobre T?

9. Seja V o espaço vetorial das funções contínuas reais definidas sobre o intervalo $[-1, 1]$ da reta real. Seja W_p o subespaço das funções pares, $f(-x) = f(x)$ e seja W_i o subespaço das funções ímpares, $f(-x) = -f(x)$.

   (a) Demonstrar que $V = W_p \oplus W_i$.

   (b) Se T é o operador integral indefinida

$$(Tf)(x) = \int_0^x f(t)\,dt$$

   W_p e W_i são invariantes sob T?

6.8 O Teorema da Decomposição Primária

Estamos tentando estudar um operador linear T sobre o espaço de V de dimensão finita, pela decomposição de T numa soma direta de operadores que sejam, num certo sentido, elementares. Podemos fazê-lo através dos valores e vetores característicos de T em certos casos particulares, isto é, quando o polinômio de T decompõe-se sobre o corpo F de escalares num produto de polinômios unitários, distintos e de grau 1. Que podemos fazer com um T arbitrário? Se tentarmos estudar T usando valores característicos, iremos nos confrontar com dois problemas. Primeiro, T poderá não ter nenhum valor característico; isto, na verdade, é uma defi-

ciência do corpo de escalares, a saber, que ele não é algebricamente fechado. Segundo, mesmo que o polinômio característico se decomponha completamente sobre F num produto de polinômios de grau 1, podem não existir vetores característicos suficientes de T para gerar o espaço V; isto é, evidentemente uma deficiência de T. A segunda situação é ilustrada pelo operador T sobre F^3 (F um corpo arbitrário) representado em relação à base canônica por

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

O polinômio característico de A é $(x-2)^2(x+1)$ e este é obviamente o polinômio minimal de A (ou de T). Assim, T não é diagonalizável. Vê-se que isto ocorre porque o núcleo de $(T-2I)$ tem dimensão 1 apenas. Por outro lado, o núcleo de $(T+I)$ e o núcleo de $(T-2I)^2$ juntos geram V, sendo o primeiro o subespaço gerado por ε_3 e o segundo o subespaço gerado por ε_1 e ε_2.

Este será mais ou menos o nosso método geral para o segundo problema. Se (lembrar que isto é uma hipótese) o polinômio minimal de T se decompõe como

$$p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}$$

sendo $c_1, \dots, c_k$ elementos distintos de F, então mostraremos que o espaço V é a soma direta dos núcleos de $(T - c_iI)^{r_i}$, $i = 1, \dots, k$. A hipótese a respeito de p é equivalente ao fato de que T seja triangulável (Teorema 5); no entanto o conhecimento deste fato não nos ajudará.

O teorema que demonstraremos é mais geral que isto que descrevemos, uma vez que trabalhará com a decomposição primária do polinômio minimal, quer sejam ou não de grau 1 os primos que comparecem na decomposição. O leitor achará útil pensar no caso particular em que os primos são de grau 1 e, de modo ainda mais particular, pensar na demonstração do Teorema 6 como um caso particular deste teorema.

Teorema 12 (Teorema da Decomposição Primária). *Seja* T *um operador linear sobre o espaço vetorial* V *de dimensão finita sobre o corpo* F. *Seja* p *o polinômio minimal de* T,

$$\mathrm{p} = \mathrm{p}_1^{r_1} \dots \mathrm{p}_k^{r_k}$$

onde os p_i *são polinômios distintos, irredutíveis e unitários sobre* F *e os* r_i *são inteiros positivos. Seja* W_i *o núcleo de* $p_i(T)^{r_i}$, $i = 1, \ldots, k$. *Então*

(i) $V = W_1 \oplus \ldots \oplus W_k$;

(ii) *cada* W_i *é invariante sob* T;

(iii) *se* T_i *é o operador induzido sobre* W_i *por* T, *então o polinômio minimal de* T_i *é* $p_i^{r_i}$.

Demonstração. A idéia da demonstração é a que segue. Se a decomposição (i) em soma direta é válida, como podemos obter as projeções $E_1, \ldots, E_k$ associadas a esta decomposição? A projeção E_i será o operador idêntico sobre W_i e zero sobre os outros W_j. Vamos determinar um polinômio h_i tal que $h_i(T)$ seja o operador idêntico sobre W_i e seja nulo sobre os outros W_j e de modo que $h_1(T) + \ldots + h_k(T) = I$, etc.

Para cada i, seja

$$f_i = \frac{p}{p_i^{r_i}} = \prod_{j \neq i} p_j^{r_j}.$$

Como $p_1, \ldots, p_k$ são polinômios primos distintos, os polinômios $f_1, \ldots, f_k$ são relativamente primos (Teorema 10, Capítulo 4). Assim, existem polinômios $g_1, \ldots, g_k$ tais que

$$\sum_{i=1}^{n} f_i g_i = 1.$$

Notemos também que se $i \neq j$, então $f_i f_j$ é divisível pelo polinômio p pois $f_i f_j$ contém cada $p_m^{r_m}$ como um fator. Vamos mostrar que os polinômios $h_i = f_i g_i$ comportam-se da maneira descrita no primeiro parágrafo da demonstração.

Seja $E_i = h_i(T) = f_i(T) g_i(T)$. Como $h_1 + \ldots + h_k = 1$ e p divide $f_i f_j$ para $i \neq j$, temos

$$E_1 + \ldots + E_k = I$$
$$E_i E_j = 0, \text{ se } i \neq j.$$

Assim, os E_i são projeções que correspondem a alguma decomposição do espaço V em soma direta. Desejamos mostrar que a imagem de E_i é exatamente o subespaço W_i. É evidente que cada vetor

na imagem de E_i está em W_i, pois se α está na imagem de E_i, então $\alpha = E_i\alpha$, logo

$$\begin{aligned} p_i(T)^{r_i}\alpha &= p_i(T)^{r_i}E_i\alpha \\ &= p_i(T)^{r_i}f_i(T)g_i(T)\alpha \\ &= 0 \end{aligned}$$

pois $p^{r_i}f_ig_i$ é divisível pelo polinômio minimal p. Reciprocamente, suponhamos que α esteja no núcleo de $p_i(T)^{r_i}$. Se $j \neq i$, então f_jg_j é divisível por $p_i^{r_i}$, logo $f_j(T)g_j(T)\alpha = 0$, isto é, $E_j\alpha = 0$ para $j \neq i$. Mas, então, é imediato que $E\alpha = \alpha$, isto é, que α está na imagem de E_i. Isto completa a demonstração da afirmação (i).

É certamente óbvio que os subespaços W_i são invariantes sob T. Se T_i é o operador induzido sobre W_i por T, então, evidentemente $p_i(T_i)^{r_i} = 0$, pois, por definição, $p_i(T)^{r_i}$ se anula no subespaço W_i. Isto mostra que o polinômio minimal de T_i divide $p_i^{r_i}$. Reciprocamente, seja g um polinômio arbitrário tal que $g(T_i) = 0$. Então $g(T)f_i(T) = 0$. Assim, gf_i é divisível pelo polinômio minimal p de T, isto é, $p_i^{r_i}f_i$ divide gf_i. Vê-se facilmente que $p_i^{r_i}$ divide g. Logo, o polinômio minimal de T_i é $p_i^{r_i}$.

Corolário. *Se* $E_1, \ldots, E_k$ *são as projeções associadas à decomposição primária de* T, *então cada* E_i *é um polinômio em* T *e, conseqüentemente, se um operador linear* U *comuta com* T *então* U *comuta com cada um dos* E_i, *isto é, cada subespaço* W_i *é invariante sob* U.

Com a notação do Teorema 12, vamos considerar rapidamente o caso particular em que o polinômio minimal de T é um produto de polinômios do primeiro grau, isto é, o caso em que cada p_i é da forma $p_i = x - c_i$. Ora, a imagem de E_i é o núcleo W_i de $(T - c_iI^{r_i})$. Coloquemos $D = c_1E_1 + \ldots + c_kE_k$. Pelo Teorema 11, D é um operador diagonalizável que denominaremos a **parte diagonalizável** de T. Consideremos o operador $N = T - D$. Ora,

$$\begin{aligned} T &= TE_1 + \ldots + TE_k \\ D &= c_1E_1 + \ldots + c_kE_k \end{aligned}$$

portanto,

$$N = (T - c_1I)E_1 + \ldots + (T - c_kI)E_k.$$

A esta altura o leitor já deverá estar suficientemente familiarizado com projeções, portanto, verá que

$$N^2 = (T - c_1 I)^2 E_1 + \dots + (T - c_k I)^2 E_k$$

e que, em geral,

$$N^r = (T - c_1 I)^r E_1 + \dots + (T - c_k I)^r E_k.$$

Quando $r \geq r_i$ para todo i, teremos $N^r = 0$, pois o operador $(T - c_i I)^r$ será então 0 sobre a imagem de E_i.

Definição. *Seja* N *um operador linear sobre o espaço vetorial* V. *Dizemos que* N *é* **nilpotente** *se existe algum inteiro positivo* r *tal que* $N^r = 0$.

Teorema 13. *Seja* T *um operador linear sobre o espaço vetorial* V *de dimensão finita sobre o corpo* F. *Suponhamos que o polinômio minimal de* T *se decomponha sobre* F *num produto de polinômios lineares. Então existe um operador diagonalizável* D *sobre* V *e um operador nilpotente* N *sobre* V *tais que*

(i) $T = D + N$;

(ii) $DN = ND$.

O operador diagonalizável D *e o operador nilpotente* N *são determinados de modo único por* (i) *e* (ii) *e cada um deles é um polinômio em* T.

Demonstração. Acabamos de observar que podemos escrever $T = D + N$ onde D é diagonalizável e N é nilpotente e também que D e N não só comutam mas também são polinômios em T. Suponhamos, agora, que também tenhamos $T = D' + N'$, sendo d' diagonalizável, N' nilpotente e $D'N' = N'D'$. Vamos demonstrar que $D = D'$ e $N = N'$.

Como D' e N' comutam entre si e $T = D' + N'$, vemos que D' e N' comutam com T. Assim, D' e N' comutam com qualquer polinômio em T, logo eles comutam com D e com N. Agora temos

$$D + N = D' + N'$$

ou

$$D - D' = N' - N$$

e todos estes quatro operadores comutam entre si. Como D e D' são ambos diagonalizáveis e comutam, eles são simultaneamente diagonalizáveis e $D - D'$ é diagonalizável. Como N e N' são ambos

nilpotentes e comutam, o operador $(N' - N)$ é nilpotente; com efeito, usando o fato de que N e N' comutam

$$(N' - N)^r = \sum_{j=0}^{r} \binom{r}{j} (N')^{r-j}(-N)^j$$

e então quando r for suficientemente grande, todos os termos nesta expressão de $(N' - N)^r$ serão nulos. (Na realidade, um operador nilpotente sobre um espaço n-dimensional deve ter sua n-ésima potência nula; no caso acima, se tomamos $r = 2n$, este número é suficientemente grande. Decorre então que $r = n$ é suficientemente grande, mas isto não é evidente a partir da expressão acima.) Ora, $D - D'$ é um operador diagonalizável que também é nilpotente. Tal operador tem de ser, obviamente, o operador nulo; de fato, sendo nilpotente, o polinômio minimal deste operador é da forma x^r para um certo $r \leq m$; mas como o operador é diagonalizável, o polinômio minimal não pode ter uma raiz múltipla, logo $r = 1$ e o polinômio minimal é simplesmente x, o que diz que o operador é nulo. Assim, vemos que $D = D'$ e $N = N'$.

Corolário. *Seja* V *um espaço vetorial de dimensão finita sobre um corpo algebricamente fechado* F, *por exemplo, o corpo dos números complexos. Então, todo operador linear* T *sobre* V *pode ser escrito como uma soma de um operador diagonalizável* D *com um operador nilpotente* N, *os quais comutam. Estes operadores* D *e* N *são únicos e são polinômios em* T.

Por esses resultados vê-se que o estudo dos operadores lineares sobre espaços vetoriais sobre um corpo algebricamente fechado fica essencialmente reduzido ao estudo dos operadores nilpotentes. Para espaços vetoriais sobre corpos não algebricamente fechados, precisamos ainda encontrar algum substituto para valores e vetores característicos. Um fato interessante é que estes dois problemas podem ser tratados simultaneamente e é isto o que faremos no próximo capítulo.

Concluindo esta seção, gostaríamos de dar um exemplo que ilustre algumas das idéias do teorema da decomposição primária. Decidimos dá-lo no final da seção porque ele usa equações diferenciais, não sendo assim álgebra linear pura.

Exemplo 14. No teorema da decomposição primária não é necessário que o espaço vetorial V tenha dimensão finita, nem é

necessário, para as partes (i) e (ii), que p seja o polinômio minimal de T. Se T é um operador linear sobre um espaço vetorial arbitrário e *se* existe um polinômio unitário p tal que $p(T) = 0$, então as partes (i) e (ii) do Teorema 12 são válidas para T com a demonstração que fizemos.

Seja n um inteiro positivo e seja V o espaço das funções n vezes continuamente diferenciáveis f sobre a reta real que satisfazem a equação diferencial

$$\text{(6-18)} \qquad \frac{d^n f}{dt^n} + a_{n-1}\frac{d^{n-1}f}{dt^{n-1}} + \ldots + a_1\frac{df}{dt} + a_0 f = 0$$

onde $a_0, \ldots, a_{n-1}$ são certos escalares fixos. Se C_n indica o espaço das funções n vezes continuamente diferenciáveis, então o espaço V das soluções desta equação diferencial é um subespaço de C_n. Se D indica o operador derivação e p é o polinômio

$$p = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

então V é o núcleo do operador $p(D)$, pois (6-18) diz simplesmente que $p(D)f = 0$. Portanto, V é invariante sob D. Consideremos agora D como um operador linear sobre o subespaço V. Então $p(D) = 0$.

Se estamos discutindo funções diferenciáveis com valores complexos, então C_n e V são espaços vetoriais complexos e $a_0, \ldots, a_{n-1}$ podem ser quaisquer números complexos. Escrevamos agora

$$p = (x - c_1)^{r_1} \ldots (x - c_k)^{r_k}$$

onde $c_1, \ldots, c_k$ são números complexos distintos. Se W_j é o núcleo de $(D - c_j I)^{r_j}$, então o Teorema 12 diz que

$$V = W_1 \oplus \ldots \oplus W_k.$$

Em outras palavras, se f satisfaz a equação diferencial (6-18), então f pode ser expressa de modo único sob a forma

$$f = f_1 + \ldots + f_k$$

onde f_j satisfaz a equação diferencial $(D - c_j I)^{r_j} f_j = 0$. Assim, o estudo das soluções da equação (6-18) fica reduzido ao estudo do espaço das soluções de uma equação diferencial da forma

(6-19) $$(D - cI)^r f = 0.$$

Esta redução foi conseguida por meio dos métodos gerais de álgebra linear, isto é, pelo teorema da decomposição primária.

Para descrever o espaço das soluções de (6-19), é necessário saber alguma coisa sobre equações diferenciais, isto é, é necessário saber alguma coisa a respeito de D além do fato de que D é um operador linear. No entanto, não se precisa saber muito. É bem fácil demonstrar por indução sobre r que, se f está em C_r, então

$$(D - cI)^r f = e^{ct} D^r (e^{-ct} f)$$

isto é,

$$\frac{df}{dt} - cf(t) = e^{ct} \frac{d}{dt}(e^{-ct} f), \text{ etc.}$$

Assim, $(D - cI)^r f = 0$ se, e somente se, $D^r(e^{-ct} f) = 0$. Uma função g tal que $D^r g = 0$, isto é, $d^r g/dt^r = 0$, deve ser uma função polinomial de grau menor ou igual a $(r - 1)$:

$$g(t) = b_0 + b_1 t + \ldots + b_{r-1} t^{r-1}.$$

Assim f satisfaz (6-19) se, somente se, f tem a forma

$$f(t) = e^{ct}(b_0 + b_1 t + \ldots + b_{r-1} t^{r-1}).$$

Conseqüentemente, as "funções" e^{ct}, te^{ct}, ..., $t^{r-1}e^{ct}$ geram o espaço das soluções de (6-19). Como 1, t, ..., t^{r-1} são funções linearmente independentes e a função exponencial não possui raízes, estas r funções $t^j e^{ct}$, $0 \leq j \leq r - 1$, formam uma base do espaço das soluções.

Voltando à equação diferencial (6-18), que é

$$p(D)f = 0$$
$$p = (x - c_1)^{r_1} \ldots (x - c_k)^{r_k}$$

vemos que as n funções $t^m e^{c_j t}$, $0 \leq m \leq r_j - 1$, $1 \leq j \leq k$, formam uma base do espaço das soluções de (6-18). Em particular, o espaço das soluções tem dimensão finita igual ao grau do polinômio p.

Exercícios

1. Seja T um operador linear sobre R^3 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{bmatrix}.$$

Exprimir o polinômio minimal p de T sob a forma $p = p_1 p_2$, sendo p_1 e p_2 unitários e irredutíveis sobre o corpo dos números reais. Seja W_i o núcleo de $p_i(T)$. Determinar bases $\mathscr{B}_i$ dos espaços W_1 e W_2. Se T_i é o operador induzido sobre W_i por T, determinar a matriz de T_i em relação à base $\mathscr{B}_i$ (acima).

2. Seja T o operador linear sobre R^3 que é representado pela matriz

$$\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$$

em relação à base ordenada canônica. Mostrar que existe um operador diagonalizável D sobre R^3 e um operador nilpotente N sobre R^3 tais que $T = D + N$ e $DN = ND$. Determinar as matrizes de D e N em relação à base canônica. (Basta repetir a demonstração do Teorema 12 para este caso particular.)

3. Se V é o espaço dos polinômios de grau menor ou igual a n sobre um corpo F, mostrar que o operador derivação sobre V é nilpotente.

4. Seja T um operador linear sobre o espaço V de dimensão finita cujo polinômio característico seja

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$$

e cujo polinômio minimal seja

$$p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}.$$

Seja W_i o núcleo de $(T - c_iI)^{r_i}$.

(a) Demonstrar que W_i é o conjunto dos vetores α em V tais que $(T - c_iI)^m\alpha = 0$ para algum inteiro positivo m (que pode depender de α)

(b) Demonstrar que a dimensão de W_i é d_i. [*Sugestão*: Se T_i é o operador induzido sobre W_i por T, então $T_i - c_iI$ é nilpotente, assim, o polinômio característico de $T_i - c_iI$ deve ser x^{e_i}, sendo e_i a dimensão de W_i (demonstração?); assim, o polinômio característico de T_i é $(x - c_i)^{e_i}$; agora usar o fato de que o polinômio característico de T é o produto dos polinômios característicos de T_i para mostrar que $e_i = d_i$.]

5. Seja V um espaço vetorial de dimensão finita sobre o corpo dos números complexos. Seja T um operador linear sobre V e seja D a parte diagonalizável de T. Demonstrar que se g é um polinômio qualquer com coeficientes complexos, então a parte diagonalizável de $g(T)$ é $g(D)$.

6. Seja V um espaço vetorial de dimensão finita sobre o corpo F e seja T um operador linear sobre V tal que posto $(T) = 1$. Demonstrar que ou T é diagonalizável ou T é nilpotente, não ambos.

7. Seja V um espaço vetorial de dimensão finita sobre F e seja T um operador linear sobre V. Suponhamos que T comute com todo operador linear diagonalizável sobre V. Demonstrar que T é um múltiplo escalar do operador idêntico.

8. Seja V o espaço das $n \times n$ matrizes sobre um corpo F e seja A uma $n \times n$ matriz fixa sobre F. Definamos um operador linear T sobre V por $T(B) = AB - BA$. Demonstrar que se A é uma matriz nilpotente, então T é um operador nilpotente.

9. Dar um exemplo de duas 4×4 matrizes nilpotentes que tenham o mesmo polinômio minimal (elas têm, necessariamente, o mesmo polinômio característico) mas que não sejam semelhantes.

10. Seja T um operador linear sobre o espaço V de dimensão finita, seja $p = p_1^{r_1}, \ldots, p_k^{r_k}$ o polinômio minimal de T e seja $V = W_1 \oplus \ldots \oplus W_k$ a decomposição primária de T, isto é, W_j é o núcleo de $p_j(T)^{r_j}$. Seja W um subespaço qualquer de V que seja invariante sob T. Demonstrar que

$$W = (W \cap W_1) \oplus (W \cap W_2) \oplus \ldots \oplus (W \cap W_k).$$

11. O que está errado na seguinte demonstração do Teorema 13? Suponhamos que o polinômio minimal de T seja um produto de fatores lineares. Então, pelo Teorema 5, T é triangulável. Seja $\mathscr{B}$ uma base ordenada tal que $A = [T]_{\mathscr{B}}$ seja triangular superior. Seja D a matriz diagonal com elementos diagonais $a_{11}, \ldots, a_{nn}$. Então $A = D + N$, onde N é estritamente triangular superior. Evidentemente N é nilpotente.

12. Se você pensou sobre o Exercício 11, pense novamente, após observar o que o Teorema 7 afirma sobre as partes diagonalizável e nilpotente de T.

13. Seja T um operador linear sobre V com polinômio minimal da forma p^n, sendo p irredutível sobre o corpo dos escalares. Mostrar que existe um vetor α em V tal que o T-anulador de α seja p^n.

14. Usar o teorema da decomposição primária e o resultado do Exercício 13 para demonstrar o seguinte: Se T é um operador linear arbitrário sobre um espaço vetorial V de dimensão finita, então existe um vetor α em V cujo T-anulador é igual ao polinômio minimal de T.

15. Se N é um operador linear nilpotente sobre um espaço vetorial n-dimensional V, então o polinômio característico de N é x^n.

Capítulo 7
AS FORMAS RACIONAL E DE JORDAN

7.1 Subespaços Cíclicos e Anuladores

Uma vez mais V é um espaço vetorial de dimensão finita sobre o corpo F e T é um operador linear fixo (mas arbitrário) sobre V. Se α é um vetor qualquer em V, existe um menor subespaço de T que é invariante sob T e contém α. Este subespaço pode ser definido como a interseção de todos os subespaços T-invariantes que contêm α; contudo, é mais proveitoso no momento considerarmos as coisas de outra maneira. Se W é um subespaço arbitrário de V que seja invariante sob T e contenha α, então W deve conter o vetor $T\alpha$; portanto, W deve conter $T(T\alpha) = T^2\alpha$, $T(T^2\alpha) = T^3\alpha$ etc. Em outras palavras, W deve conter $g(T)\alpha$ para todo polinômio g sobre F. O conjunto dos vetores da forma $g(T)\alpha$, com g em $F[x]$ é evidentemente invariante sob T e é assim o menor subespaço T-invariante que contém α.

Definição. *Se* α *é um vetor qualquer em* V, *o* **subespaço T-cíclico gerado por** α *é o subespaço* $Z(\alpha; T)$ *dos vetores da forma* $g(T)\alpha$, g *em* F[x]. *Se* $Z(\alpha; T) = V$, *então* α *é denominado um* **vetor cíclico** *de* T.

Outra maneira de descrever o subespaço $Z(\alpha; T)$ é dizer que $Z(\alpha; T)$ é o subespaço gerado pelos vetores $T^k\alpha$, $k \geq 0$, e assim α é um vetor cíclico de T se, e somente se, estes vetores geram V. Prevenimos o leitor de que um operador genérico T não possui vetores cíclicos.

Exemplo 1. Para T arbitrário, o subespaço T-cíclico gerado pelo vetor nulo é o subespaço nulo. O espaço $Z(\alpha; T)$ é unidimensional se, e somente se, α é um vetor característico de T. Para o operador idêntico, todo vetor não-nulo gera um subespaço cíclico unidimensional; assim, se $\dim V > 1$, o operador idêntico não

possui nenhum vetor cíclico. Um exemplo de um operador que possui um vetor cíclico é o operador linear T sobre F^2 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Neste caso o vetor cíclico (um vetor cíclico) é ε_1; de fato, se $\beta = (a, b)$, então, para $g = a + bx$, temos $\beta = g(T)\varepsilon_1$. Para este mesmo operador T, o subespaço cíclico gerado por ε_2 é o espaço unidimensional gerado por ε_2, porque ε_2 é um vetor característico de T.

Para quaisquer T e α, estaremos interessados em relações lineares

$$c_0\alpha + c_1 T\alpha + \ldots + c_k T^k\alpha = 0$$

entre os vetores $T^j\alpha$, isto é, estaremos interessados nos polinômios $g = c_0 + c_1 x + \ldots + c_k x^k$ que tenham a propriedade de que $g(T)\alpha = 0$. O conjunto dos g em $F[x]$ tais que $g(T)\alpha = 0$ é evidentemente um ideal em $F[x]$. É também um ideal não-nulo, pois contém o polinômio minimal p do operador T ($p(T)\alpha = 0$ para todo α em V).

Definição. *Se* α *é um vetor arbitrário em* V, *o* T-**anulador** *de* α *é o ideal* $M(\alpha; T)$ *em* $F[x]$ *formado pelos polinômios* g *sobre* F *tais que* $g(T)\alpha = 0$. *O único polinômio unitário* p_α *que gera este ideal também será denominado o* T-**anulador** *de* α.

Como ressaltamos acima, o T-anulador p_α divide o polinômio minimal do operador T. O leitor deverá notar também que $\text{gr}(p_\alpha) > 0$ a não ser quando α é o vetor nulo.

Teorema 1. *Seja* α *um vetor não nulo arbitrário em* V *e seja* p_α *o* T-*anulador de* α.

(i) *O grau de* p_α *é igual à dimensão do subespaço cíclico* $Z(\alpha; T)$.

(ii) *Se o grau de* p_α *é* k, *então os vetores* $\alpha, T\alpha, T^2\alpha, \ldots, T^{k-1}\alpha$ *formam uma base de* $Z(\alpha; T)$.

(iii) *Se* U *é o operador linear sobre* $Z(\alpha; T)$ *induzido por* T, *então o polinômio minimal de* U *é* p_α.

Demonstração. Seja g um polinômio qualquer sobre o corpo F. Podemos escrever

$$g = p_\alpha q + r$$

onde $r = 0$ ou $\text{gr}(r) < \text{gr}(p_\alpha) = k$. O polinômio $p_\alpha q$ está no T-anulador de α, portanto

$$g(T)\alpha = r(T)\alpha.$$

Como $r = 0$ ou $\text{gr}(r) < k$, o vetor $r(T)\alpha$ é uma combinação linear dos vetores α, $T\alpha$, ..., $T^{k-1}\alpha$ e como $g(T)\alpha$ é um vetor típico em $Z(\alpha; T)$ isto mostra que estes k vetores geram $Z(\alpha; T)$. Estes vetores são, sem dúvida, linearmente independentes, pois toda relação linear não-trivial entre eles nos forneceria um polinômio não-nulo g tal que $g(T)\alpha = 0$ e $\text{gr}(g) < \text{gr}(p_\alpha)$, o que é absurdo. Isto demonstra (i) e (ii).

Seja U o operador linear sobre $Z(\alpha; T)$ obtido pela restrição de T àquele subespaço. Se g é um polinômio arbitrário sobre F, então

$$\begin{aligned} p_\alpha(U)\, g(T)\alpha &= p_\alpha(T) g(T)\alpha \\ &= g(T) p_\alpha(T)\alpha \\ &= g(T)0 \\ &= 0. \end{aligned}$$

Assim, o operador $p_\alpha(U)$ leva todo vetor de $Z(\alpha; T)$ em 0 e é, portanto, o operador nulo sobre $Z(\alpha; T)$. Além disso, se h é um polinômio de grau menor que k, não podemos ter $h(U) = 0$, pois então $h(U)\alpha = h(T)\alpha = 0$, contrariando a definição de p_α. Isto mostra que p_α é o polinômio minimal de U.

Uma conseqüência particular deste teorema é a seguinte: Se acontecer que α seja um vetor cíclico de T, então o polinômio minimal de T deve ter grau igual à dimensão do espaço V; daqui decorre, pelo teorema de Cayley-Hamilton, que o polinômio minimal de T é o polinômio característico de T. Demonstraremos posteriormente que para todo T existe um vetor α em V cujo anulador é o polinômio minimal de T. Decorrerá então que T possui um vetor cíclico se, e somente se, os polinômios minimal e característico de T são idênticos. Mas teremos algum trabalho para chegarmos a ver isto.

Nosso plano é estudar um T arbitrário usando operadores que possuam um vetor cíclico. Portanto, consideremos um operador linear U sobre um espaço W de dimensão k que possua um vetor cíclico α. Pelo Teorema 1, os vetores α, ..., $U^{k-1}\alpha$ formam uma

base do espaço W e o anulador p_α de α é o polinômio minimal de U (logo, é também o polinômio característico de U). Se fizermos $\alpha_i = U^{i-1}\alpha$, $i = 1, \ldots, k$, então a ação de U sobre a base ordenada $\mathcal{B} = \{\alpha_1, \ldots, \alpha_k\}$ é

$$(7\text{-}1)\qquad \begin{aligned} U\alpha_i &= \alpha_{i+1}, \quad i = 1, \ldots, k-1 \\ U\alpha_k &= -c_0\alpha_1 - c_1\alpha_2 - \ldots - c_{k-1}\alpha_k \end{aligned}$$

onde $p_\alpha = c_0 + c_1x + \ldots + c_{k-1}x^{k-1} + x^k$. A expressão de $U\alpha_k$ decorre do fato de que $p_\alpha(U)\alpha = 0$, isto é,

$$U^k\alpha + c_{k-1}U^{k-1}\alpha + \ldots + c_1U\alpha + c_0\alpha = 0.$$

Isto diz que a matriz de U em relação à base ordenada $\mathcal{B}$ é

$$(7\text{-}2)\qquad \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & 0 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -c_{k-1} \end{bmatrix}.$$

A matriz (7-2) é denominada a **matriz associada** ao polinômio unitário p_α.

Teorema 2. *Se* U *é um operador linear sobre o espaço de* W *dimensão finita, então* U *possui um vetor cíclico se, e somente se, existe alguma base ordenada de* W *em relação à qual* U *é representado pela matriz associada ao polinômio minimal de* U.

Demonstração. Acabamos de observar que se U possui um vetor cíclico, então existe uma tal base ordenada de W. Reciprocamente, se existe uma base ordenada $\{\alpha_1, \ldots, \alpha_k\}$ de W em relação à qual U é representado pela matriz associada ao seu polinômio minimal, é óbvio que α_1 é um vetor cíclico de U.

Corolário. *Se* A *é a matriz associada a um polinômio unitário* p, *então* p *é tanto o polinômio minimal como o polinômio característico de* A.

Demonstração. Uma maneira de ver isso é indicar por U o operador linear sobre F^k representado pela matriz A, em relação à base ordenada canônica e aplicar o Teorema 1 juntamente com

o teorema de Cayley-Hamilton. Um outro método é usar o Teorema 1 para ver que p é o polinômio minimal de A e verificar, por um cálculo direto, que p é o polinômio característico de A.

Um último comentário — se T é um operador linear arbitrário sobre o espaço V e α é um vetor qualquer em V, então o operador U que T induz sobre o subespaço cíclico $Z(\alpha; T)$ possui um vetor cíclico, a saber, α. Assim, $Z(\alpha; T)$ possui uma base ordenada em relação à qual U é representado pela matriz associada a p_α, sendo p_α o T-anulador de α.

Exercícios

1. Seja T um operador linear sobre F^2. Demonstrar que todo vetor não-nulo que não seja um vetor característico de T é um vetor cíclico de T. Depois, demonstrar que ou T possui um vetor cíclico ou então T é um múltiplo escalar do operador idêntico.

2. Seja T o operador linear sobre R^3 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Demonstrar que T não possui vetores cíclicos. Qual é o subespaço T-cíclico gerado pelo vetor $(1, -1, 3)$?

3. Seja T o operador linear sobre C^3 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} 1 & i & 0 \\ -1 & 2 & -i \\ 0 & 1 & 1 \end{bmatrix}.$$

Determinar o T-anulador do vetor $(1, 0, 0)$. Determinar o T-anulador de $(1, 0, i)$.

4. Demonstrar que se T^2 possui um vetor cíclico, então T possui um vetor cíclico. A recíproca é verdadeira?

5. Seja V um espaço vetorial n-dimensional sobre o corpo F e seja N um operador linear nilpotente sobre V. Suponhamos que $N^{n-1} \neq 0$ e seja α um vetor arbitrário em V tal que $N^{n-1}\alpha \neq 0$. Demonstrar que α é um vetor cíclico de N. Dizer exatamente qual é a matriz de N em relação à base ordenada $\{\alpha, N\alpha, \ldots, N^{n-1}\alpha\}$?

6. Demonstrar diretamente que se A é a matriz associada ao polinômio unitário p, então p é o polinômio característico de A.

7. Seja V um espaço vetorial n-dimensional e seja T um operador linear sobre V. Suponhamos que T seja *diagonalizável*.

(a) Se T possui um vetor cíclico, mostrar que T possui n-valores característicos distintos.

(b) Se T possui n valores característicos distintos e se $\{\alpha_1, \ldots, \alpha_n\}$ é uma base formada de vetores característicos de T, mostrar que $\alpha = \alpha_1, + \ldots + \alpha_n$ é um vetor cíclico de T.

8. Seja T um operador linear sobre o espaço vetorial V de dimensão finita. Suponhamos que T possua um vetor cíclico. Demonstrar que se U é um operador linear arbitrário que comuta com T, então U é um polinômio em T.

7.2 Decomposições Cíclicas e a Forma Racional

O objetivo primordial desta seção é demonstrar que se T é um operador linear qualquer sobre V, então existem vetores $\alpha_1, \ldots, \alpha_r$ em V tais que

$$V = Z(\alpha_1; T) \oplus \ldots \oplus Z(\alpha_r; T).$$

Em outras palavras, desejamos demonstrar que V é a soma direta de subespaços T-cíclicos. Isto nos mostrará que T é a soma direta de um número finito de operadores lineares, cada um dos quais possui um vetor cíclico. O resultado disto será a redução de muitos problemas relativos a um operador genérico a problemas semelhantes sobre um operador que possua um vetor cíclico. O teorema que vamos demonstrar (Teorema 3) é um dos resultados mais profundos da álgebra linear e possui muitos corolários interessantes.

O teorema da decomposição cíclica está intimamente relacionado com a seguinte questão: quais subespaços W T-invariantes possuem a propriedade de que exista um subespaço W' T-invariante tal que $V = W \oplus W'$? Se W é um subespaço arbitrário de um espaço V de dimensão finita, então existe um subespaço W' tal que $V = W \oplus W'$. Usualmente existem muitos desses subespaços W' e cada um deles é denominado um suplementar de W. Estamos perguntando quando é que um subespaço T-invariante possui um subespaço suplementar que também seja invariante sob T.

Suponhamos que $V = W \oplus W'$, onde W e W' são ambos invariantes sob T e vejamos o que é possível dizer a respeito do subespaço W. Cada vetor β em V é da forma $\beta = \gamma + \gamma'$, com γ em W e γ' em W'. Se f é um polinômio sobre o corpo de escalares, então

$$f(T)\beta = f(T)\gamma + f(T)\gamma'.$$

Como W e W' são invariantes sob T, o vetor $f(T)\gamma$ está em W e $f(T)\gamma'$ está em W'. Portanto, $f(T)\beta$ está em W se, e somente se, $f(T)\gamma' = 0$. O que nos interessa é o fato, aparentemente inocente, de que se $f(T)\beta$ está em W, então $f(T)\beta = f(T)\gamma$.

Definição. *Seja* T *um operador linear sobre um espaço vetorial* V *e seja* W *um subespaço de* V. *Dizemos que* W *é* T-**admissível** *se*

(i) W *é invariante sob* T;

(ii) *se* f(T)β *está em* W, *existe um vetor* γ *em* W *tal que* f(T)β = f(T)γ.

Como acabamos de mostrar, se W for invariante e tiver um subespaço suplementar invariante, então W será admissível. Uma das conseqüências do Teorema 3 será a recíproca dessa afirmação, de modo que a propriedade de ser admissível caracterizará aqueles subespaços invariantes que possuam subespaços suplementares invariantes.

Vejamos como a propriedade de ser admissível se envolve na tentativa de obtermos uma decomposição

$$V = Z(\alpha_1; T) \oplus \dots \oplus Z(\alpha_r; T).$$

O processo básico para chegarmos a uma tal decomposição será o de selecionar indutivamente os vetores $\alpha_1, \dots, \alpha_r$. Suponhamos que por um processo ou outro tenhamos escolhido $\alpha_1, \dots, \alpha_j$ e que o subespaço

$$W_j = Z(\alpha_1; T) + \dots + Z(\alpha_j; T)$$

seja próprio. Gostaríamos de encontrar um vetor não-nulo α_{j+1} tal que

$$W_j \cap Z(\alpha_{j+1}; T) = \{0\}$$

pois nesse caso o subespaço $W_{j+1} = W_j \oplus Z(\alpha_{j+1}; T)$ se aproximará, de pelo menos mais uma dimensão, ao ponto de esgotar V. Mas por que deverá existir um α_{j+1} nessas condições? Se $\alpha_1, \dots, \alpha_j$ foram escolhidos de modo que W_j seja um subespaço T-admissível, é bastante fácil ver que podemos encontrar um α_{j+1} conveniente. É esse fato que fará funcionar nossa demonstração do Teorema 3, apesar de não ser assim que iremos formular o argumento.

Seja W um subespaço próprio T-invariante. Tentemos encontrar um vetor não-nulo α tal que

(7-3) $$W \cap Z(\alpha; T) = \{0\}.$$

Podemos escolher um vetor β que não esteja em W. Consideremos o *T-condutor* $S(\beta; W)$, formado de todos os polinômios g tais que $g(T)\beta$ esteja em W. Recordemos que o polinômio unitário $f = s(\beta; W)$, que gera o ideal $S(\beta; W)$ também é denominado o *T-condutor de* β *em* W. O vetor $f(T)\beta$ está em W. Ora, se W é T-admissível, existe um γ em W com $f(T)\beta = f(T)\gamma$. Seja $\alpha = \beta - \gamma$ e seja g um polinômio arbitrário. Como $\beta - \alpha$ está em W, $g(T)\beta$ estará em W se, e somente se, $g(T)\alpha$ estiver em W; em outras palavras, $S(\alpha; W) = S(\beta; W)$. Portanto, o polinômio f é também um T-condutor de α em W. Mas $f(T)\alpha = 0$. Isso nos diz que $g(T)\alpha$ estará em W se, e somente se, $g(T)\alpha = 0$, isto é, os subespaços $Z(\alpha; T)$ e W são independentes (7-3) e f é o T-anulador de α.

Teorema 3 (Teorema da Decomposição Cíclica). *Seja* T *um operador linear sobre um espaço vetorial* V *de dimensão finita e seja* W_0 *um subespaço próprio* T-*admissível de* V. *Então existem vetores não-nulos* $\alpha_1, \ldots, \alpha_r$ *em* V *com respectivos* T-*anuladores* $p_1, \ldots, p_r$ *tais que*

(i) $V = W_0 \oplus Z(\alpha_1; T) \oplus \ldots \oplus Z(\alpha_r; T)$;

(ii) p_k *divide* p_{k-1}, $k = 2, \ldots, r$.

Além disso, o inteiro r *e os anuladores* $p_1, \ldots, p_r$ *são determinados de modo único por* (i), (ii) *e pelo fato de que nenhum* α_k *é nulo.*

Demonstração. A demonstração é bastante longa; por isso nós a dividiremos em quatro partes. Numa primeira leitura será mais simples tomarmos $W_0 = \{0\}$ apesar de que esse fato não produza nenhuma simplificação substancial. Ao longo da demonstração escreveremos $f\beta$ no lugar de $f(T)\beta$.

Primeira parte. Existem vetores não-nulos $\beta_1, \ldots, \beta_r$ *em* V *tais que*

(a) $V = W_0 + Z(\beta_1; T) + \ldots + Z(\beta_r; T)$;

(*a*) *se* $1 \le k \le r$ *e*

$$W_k = W_0 + Z(\beta_1; T) + \ldots + Z(\beta_k; T),$$

o condutor $p_k = s(\beta_k; W_{k-1})$ *tem grau máximo entre todos os* T-*condutores no espaço* W_{k-1}, *isto é, para cada* k

$$gr(p_k) = \max_{\alpha \text{ em } V} gr(s(\alpha; W_{k-1})).$$

Esta parte depende somente do fato de que W_0 é um subespaço invariante. Se W é um subespaço próprio T-invariante, então

$$0 < \max_{\alpha} \operatorname{gr}(s(\alpha; W)) \leq \dim V$$

e podemos escolher um vetor β de modo que $\operatorname{gr}(s(\beta; W))$ atinja esse máximo. O subespaço $W + Z(\beta; T)$ é então T-invariante e possui dimensão maior que $\dim W$. Apliquemos esse processo a $W = W_0$ para obter β_1. Se $W_1 = W_0 + Z(\beta_1; T)$ é ainda próprio, então apliquemos o processo a W_1 para obter β_2. Continuemos dessa maneira. Como $\dim W_k > \dim W_{k-1}$, devemos obter $W_r = V$ em, no máximo, $\dim V$ passagens.

Segunda parte. Sejam $\beta_1, \ldots, \beta_r$ *vetores não-nulos satisfazendo as condições* (a) *e* (b) *da primeira parte. Fixemos* k, $1 \leq k \leq r$. *Seja* β *um vetor arbitrário de* V *e seja* $f = s(\beta; W_{k-1})$. *Se*

$$f\beta = \beta_0 + \sum_{1 \leq i \leq k} g_i \beta_i, \qquad \beta_i \textit{ em } W_i$$

então f *divide cada polinômio* g_i *e* $\beta_0 = f\gamma_0$ *com* γ_0 *em* W_0.

Se $k = 1$ estamos afirmando exatamente que W_0 é *T-admissível.* Para provar a afirmação para $k > 1$, apliquemos o algoritmo da divisão:

$$g_i = fh_i + r_i, \qquad r_i = 0 \quad \text{ou} \quad \operatorname{gr} r_i < \operatorname{gr} f. \tag{7-4}$$

Queremos mostrar que $r_i = 0$ para cada i. Seja

$$\gamma = \beta - \sum_{1}^{k-1} h_i \beta_i \tag{7-5}$$

Como $\gamma - \beta$ está em W_{k-1},

$$s(\gamma; W_{k-1}) = s(\beta; W_{k-1}) = f.$$

Além do mais

$$f\gamma = \beta_0 + \sum_{1}^{k-1} r_i \beta_i. \tag{7-6}$$

Suponhamos que algum r_i seja diferente de zero. Mostraremos que isso nos conduzirá a uma contradição. Seja j o maior índice i para o qual $r_i \neq 0$. Então

(7-7) $$f\gamma = \beta_0 + \sum_1^j r_i\beta_i, \quad r_j \neq 0 \quad \text{e} \quad \text{gr}\, r_j < \text{gr}\, f.$$

Seja $p = s(\gamma; W_{j-1})$. Como W_{k-1} contém W_{j-1}, o condutor $f = s(\gamma; W_{k-1})$ necessariamente divide p:

$$p = fg.$$

Apliquemos $g(T)$ a ambos os membros de (7-7):

(7-8) $$p\gamma = gf\gamma = gr_j\beta_j + g\beta_0 + \sum_{1 \leq i < j} gr_i\beta_i.$$

Por definição, $p\gamma$ está em W_{j-1} e os dois últimos termos no segundo membro de (7-8) estão em W_{j-1}. Portanto, $\text{gr}_j\beta_j$ está em W_{j-1}. Usemos agora a condição (b) da primeira parte:

$$\begin{aligned} \text{gr}\,(gr_j) &\geq \text{gr}\, s(\beta_j; W_{j-1}) \\ &= \text{gr}\, p_j \\ &\geq \text{gr}\, s(\gamma; W_{j-1}) \\ &\text{gr}\, p \\ &\text{gr}\,(fg). \end{aligned}$$

Assim, $\text{gr}(r_j) \geq \text{gr}(f)$ e isso contradiz a escolha de j. Sabemos agora que f divide cada g_i e, portanto, que $\beta_0 = \text{f}\gamma$. Como W_0 é T-admissível, $\beta_0 = f\gamma_0$ com γ_0 em W_0. Observemos, de passagem, que a segunda parte é uma forma mais forte da afirmação de que cada um dos subespaços $W_1, W_2, \ldots, W_r$ é T-admissível.

Terceira parte. Existem vetores não-nulos $\alpha_1, \ldots, \alpha_r$ *em* **V** *que satisfazem as condições* (i) *e* (ii) *do Teorema* 3.

Comecemos com os vetores $\beta_1, \ldots, \beta_r$ como na primeira parte. Fixemos k, $1 \leq k \leq r$. Apliquemos o resultado da segunda parte ao vetor $\beta = \beta_k$ e ao T-condutor $f = p_k$. Obteremos

(7-9) $$p_k\beta_k = p_k\gamma_0 + \sum_{1 \leq i < k} p_k h_i \beta_i$$

onde γ_0 está em W_0 e $h_1, \ldots, h_{k-1}$ são polinômios. Seja

(7-10) $$\alpha_k = \beta_k - \gamma_0 - \sum_{1 \leq i < k} h_i\beta_i.$$

Como $\beta_k - \alpha_k$ está em W_{k-1},

(7-11) $$s(\alpha_k;\ W_{k-1}) = s(\beta_k;\ W_{k-1}) = p_k$$

e como $p_k\alpha_k = 0$, temos

(7-12) $$W_{k-1} \cap Z(\alpha_k;\ T) = \{0\}.$$

Cada α_k satisfaz (7-11) e (7-12). Segue-se que

$$W_k = W_0 \oplus Z(\alpha_1;\ T) \oplus \ldots \oplus Z(\alpha_k;\ T)$$

e p_k é o T-anulador de α_k. Em outras palavras, os vetores $\alpha_1, \ldots, \alpha_r$ definem a mesma seqüência de subespaços W_1, W_2, ... que os vetores $\beta_1, \ldots, \beta_r$ e os T-condutores $p_k = s(\alpha_k, W_{k-1})$ possuem as mesmas propriedades de maximalidade (condição (b) da primeira parte). Os vetores $\alpha_1, \ldots, \alpha_r$ possuem a propriedade adicional de que os subespaços W_0, $Z(\alpha_1;\ T)$, $Z(\alpha_2;\ T)$, ... são independentes. É portanto fácil verificar a condição (ii) do Teorema 3. Como $p_i\alpha_i = 0$ para cada i, temos as relações triviais

$$p_k\alpha_k = 0 + p_1\alpha_1 + \ldots + p_{k-1}\alpha_{k-1}.$$

Apliquemos o resultado da segunda parte, com $\alpha_1, \ldots, \alpha_k$ no lugar de $\beta_1, \ldots, \beta_k$ e $\beta = \alpha_k$. Conclusão: p_k divide cada p_i com $i < k$.

Quarta parte. O número r *e os polinômios* $\mathrm{p}_1, \ldots, \mathrm{p}_r$ *são determinados de modo único pelas condições do Teorema* 3.

Suponhamos que além dos vetores $\alpha_1, \ldots, \alpha_r$ do Teorema 3, tenhamos os vetores não-nulos $\gamma_1, \ldots, \gamma_s$, com respectivos T-anuladores $g_1, \ldots, g_s$, tais que

(7-13) $$\begin{aligned} &V = W_0 \oplus Z(\gamma_1;\ T) \oplus \ldots \oplus Z(\gamma_s;\ T) \\ &g_k \text{ divide } g_{k-1}, \qquad k = 2, \ldots, s. \end{aligned}$$

Mostraremos que $r = s$ e $p_i = g_i$ para cada i.

É muito fácil ver que $p_1 = g_1$. O polinômio g_1 é determinado a partir de (7-13), como o T-condutor de V em W_0. Seja $S(V;\ W_0)$ a coleção dos polinômios f tais que $f\beta$ esteja em W_0 para todo β em V, isto é, polinômios f tais que a imagem de $f(T)$ esteja contida em W_0. Então $S(V; W_0)$ será um ideal não nulo da álgebra

dos polinômios. O polinômio g_1 é, pela mesma razão, o gerador unitário desse ideal. Cada β em V tem a forma

$$\beta = \beta_0 + f_1\gamma_1 + \ldots + f_s\gamma_s$$

e, assim,

$$g_1\beta = g_1\beta_0 + \sum_{1}^{s} g_1 f_i \gamma_i.$$

Como cada g_i divide g_1, temos $g_1\gamma_i = 0$ para todo i e $g_1\beta = g_1\beta_0$ está em W_0. Logo, g_1 está em $S(V; W_0)$. Como g_1 é o polinômio unitário de menor grau que leva γ_1 em W_0, vemos que g_1 é o polinômio unitário de menor grau no ideal $S(V; W_0)$. Pelo mesmo argumento, p_1 é o gerador desse ideal, logo $p_1 = g_1$.

Se f é um polinômio e W um subespaço de V, indicaremos abreviadamente por fW o conjunto de todos os vetores $f\alpha$ com α em W. Deixaremos como exercício a demonstração dos seguintes três fatos:

1. $fZ(\alpha; T) = Z(f\alpha; T)$.
2. Se $V = V_1 \oplus \ldots \oplus V_k$, onde cada V_i é invariante sob T, então $fV = fV_1 \oplus \ldots \oplus fV_k$.
3. Se α e γ tiverem o mesmo T-anulador, então $f\alpha$ e $f\gamma$ terão o mesmo T-anulador e (portanto)

$$\dim Z(f\alpha; T) = \dim Z(f\gamma; T).$$

Procederemos, agora, por indução, mostrando que $r = s$ e $p_i = g_i$ para $i = 2, \ldots, r$. O argumento consistirá em contar dimensões de maneira apropriada. Demonstraremos que se $r \geq 2$, então $p_2 = g_2$ e a partir disso, a indução tornar-se-á clara. Suponhamos $r \geq 2$. Então

$$\dim W_0 + \dim Z(\alpha_1; T) < \dim V.$$

Como sabemos que $p_1 = g_1$, saberemos também que $Z(\alpha_1; T)$ e $Z(\gamma_1; T)$ têm a mesma dimensão. Portanto,

$$\dim W_0 + \dim Z(\gamma_1; T) < \dim V$$

mostrando que $s \geq 2$. Agora tem sentido perguntarmos se $p_2 = g_2$. Das duas decomposições de V, obtemos duas decomposições do subespaço p_2V:

(7-14)
$$p_2V = p_2W_0 \oplus Z(p_2\alpha_1; T)$$
$$p_2V = p_2W_0 \oplus Z(p_2\gamma_1; T) \oplus \ldots \oplus Z(p_2\gamma_s; T).$$

Fizemos uso dos fatos (1) e (2) acima e também do fato de que $p_2\alpha_i = 0$, $i \geq 2$. Como sabemos que $p_1 = g_1$, o fato (3) acima nos diz que $Z(p_2\alpha_1; T)$ e $Z(p_2\gamma_1; T)$ têm a mesma dimensão. Portanto, usando (7-14) vemos que

$$\dim Z(p_2\gamma_i; T) = 0, \qquad i \geq 2.$$

Concluímos que $p_2\gamma_2 = 0$ e g_2 divide p_2. O argumento pode ser invertido para mostrar que p_2 divide g_2. Portanto, $p_2 = g_2$.

Corolário. *Se* T *é um operador linear sobre um espaço vetorial de dimensão finita, então todo subespaço* T*-admissível possui um subespaço suplementar que também é invariante sob* T.

Demonstração. Seja W_0 um subespaço admissível de V. Se $W_0 = V$, o suplemento que procuramos é $\{0\}$. Se W_0 é próprio, Apliquemos o Teorema 3 e seja

$$W'_0 = Z(\alpha_1; T) \oplus \ldots \oplus Z(\alpha_r; T).$$

Então W'_0 é invariante sob T e $V = W_0 \oplus W'_0$.

Corolário. *Seja* T *um operador linear sobre um espaço vetorial* V *de dimensão finita. Então*:

(a) *Existe um vetor* α *em* V *tal que o* T*-anulador de* α *seja o polinômio minimal de* T.

(b) T *possui um vetor cíclico se, e somente se, os polinômios característico e minimal de* T *são idênticos.*

Demonstração. Se $V = \{0\}$, os resultados são trivialmente verdadeiros. Se $V \neq \{0\}$, seja

(7-15)
$$V = Z(\alpha_1; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

onde os T-anuladores $p_1, \ldots, p_r$ são tais que p_{k+1} divida p_k, $1 \leq k \leq r-1$. Como observamos na demonstração do Teorema 3, segue-se facilmente que p_i é o polinômio minimal de T, isto é, o T-condutor de V em $\{0\}$. Assim demonstramos (a).

Vimos na Seção 7.1 que, se T possui um vetor cíclico, o polinômio minimal de T coincide com o polinômio característico. O conteúdo de (b) está na recíproca. Escolhamos um α arbitrário como em (1). Se o grau do polinômio minimal for dim V, então $V = Z(\alpha; T)$.

Teorema 4. (Teorema generalizado de Cayley-Hamilton). *Seja* T *um operador linear sobre um espaço vetorial* V *de dimensão finita. Sejam* p *e* f, *respectivamente, os polinômios minimal e característico de* T. *Então*

(i) p *divide* f.

(ii) p *e* f *possuem os mesmos fatores primos a menos de multiplicidades.*

(iii) *Se*

(7-16) $$p = f_1^{r_1} \dots f_k^{r_k}$$

é a decomposição em fatores primos de p, *então*

(7-17) $$f = f_1^{d_1} \dots f_k^{d_k}$$

onde d_i *é a nulidade de* $f_i(T)^{r_i}$ *dividida pelo grau de* f_i.

Demonstração. Ignoraremos o caso trivial $V = \{0\}$. Para demonstrar (i) e (ii) consideremos a decomposição cíclica (7-15) de V, obtida do Teorema 3. Como observamos na demonstração do segundo corolário, $p_1 = p$. Seja U_i a restrição de T a $Z(\alpha_i; T)$. Então U_i possui um vetor cíclico e assim p_i é tanto o polinômio minimal como o polinômio característico de U_i. Portanto, o polinômio característico f é o produto $f = p_1 \dots p_r$. Isso é evidente a partir da forma em blocos (6-14) que a matriz de T assume em relação a uma base conveniente. Obviamente, $p_1 = p$ divide f e isso demonstra (i). É claro que qualquer divisor primo de p é um divisor primo de f. Reciprocamente, um divisor primo de $f = p_1 \dots p_r$ divide, necessariamente, um dos fatores p_i, que por sua vez divide p_1.

Seja (7-16) a decomposição em fatores primos de p. Apliquemos o teorema da decomposição primária (Teorema 12 do Capítulo 6). Ele afirma que se V_i é o núcleo de $f_i(T)^{r_i}$, então

(7-18) $$V = V_1 \oplus \dots \oplus V_k$$

e $f_i^{r_i}$ é o polinômio minimal do operador T_i, obtido pela restrição de T ao subespaço (invariante) V_i. Apliquemos a parte (ii) deste teorema ao operador T_i. Como seu polinômio minimal é uma potência do polinômio primo f_i, o polinômio característico de T_i tem a forma $f_i^{d_i}$, com $d_i \geq r_i$. Obviamente

$$d_i = \frac{\dim V_i}{\text{gr}\,(f_i)}$$

e (quase por definição) $\dim V_i = \text{nulidade}\, f_i(T)^{r_i}$. Como T é a soma direta dos operadores $T_1, \ldots, T_k$, o polinômio característico f é o produto

$$f = f_1^{d_1} \ldots f_k^{d_k}.$$

Corolário. *Se* T *é um operador linear nilpotente sobre um espaço vetorial de dimensão* n, *o polinômio característico de* T *é* x^n.

Observamos agora o análogo matricial do teorema de decomposição cíclica. Se temos o operador T e a decomposição em soma direta do Teorema 3, seja $\mathscr{B}_i$ a "base ordenada cíclica"

$$\{\alpha_i, T\alpha_i, \ldots, T^{k_i-1}\alpha_i\}$$

de $Z(\alpha_i; T)$. Aqui k_i indica a dimensão de $Z(\alpha_i; T)$, ou seja, o grau do anulador p_i. A matriz do operador induzido T_i em relação à base ordenada $\mathscr{B}_i$ é a matriz associada ao polinômio p_i. Assim, se $\mathscr{B}$ é a base ordenada de V obtida pela reunião das $\mathscr{B}_i$, ordenadas como $\mathscr{B}_1, \ldots, \mathscr{B}_r$, então a matriz de T em relação à base ordenada $\mathscr{B}$ será

$$A = \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \ldots & A_r \end{bmatrix} \tag{7-19}$$

onde A_i é a $k_i \times k_i$ matriz associada a p_i. Diremos que uma $n \times n$ matriz A, que seja a soma direta (7-19) das matrizes associadas a polinômios não-constantes e unitários $p_1, \ldots, p_r$ tais que p_{i+1} divide p_i para $i = 1, \ldots, r-1$, está sob a **forma racional.** O teorema da decomposição cíclica nos diz o seguinte em relação a matrizes:

Teorema 5. *Seja* F *um corpo e seja* B *uma* n × n *matriz sobre* F. *Então* B *é semelhante sobre o corpo* F *a uma e somente a uma matriz sob a forma racional.*

Demonstração. Seja T o operador linear sobre F^n que é representado por B em relação à base ordenada canônica. Como acabamos de observar, existe alguma base ordenada de F^n em relação à qual T é representado por uma matriz A sob a forma racional. Então B é semelhante a esta matriz A. Suponhamos que B seja semelhante sobre F a uma outra matriz C que esteja sob a forma racional. Isto significa simplesmente que existe alguma base ordenada de F^n em relação à qual o operador T é representado pela matriz C. Se C é a soma direta das matrizes C_i associadas a polinômios unitários $g_1, \ldots, g_s$ tais que g_{i+1} divida g_i para $i = 1, \ldots, s-1$ então é evidente que teremos vetores não-nulos $\beta_1, \ldots, \beta_s$ em V com T-anuladores $g_1, \ldots, g_s$ tais que

$$V = Z(\beta_1; T) \oplus \ldots \oplus Z(\beta_s; T).$$

Mas então, pela afirmação da unicidade no teorema da decomposição cíclica, os polinômios g_i são idênticos aos polinômios p_i que definem a matriz A. Assim, $C = A$.

Os polinômios $p_1, \ldots, p_r$ são denominados **fatores invariantes** da matriz B. Na Seção 7.4 descreveremos um algoritmo que permite calcular os fatores invariantes de uma matriz B dada. O fato de que é possível calcular esses polinômios por meio de um número finito de operações racionais sobre os elementos de B, é que dá às formas racionais o seu nome.

Exemplo 2. Suponhamos que V seja um espaço vetorial bidimensional sobre o corpo F e T seja um operador linear sobre V. As possibilidades para a decomposição de T em subespaços cíclicos bastante limitadas. De fato, se o polinômio minimal de T tem grau 2, ele é igual ao polinômio característico de T e T possui um vetor cíclico. Assim, existe uma base ordenada de V em relação à qual T é representado pela matriz associada ao seu polinômio característico. Se, por outro lado, o polinômio minimal de T tem grau 1, então T é um múltiplo escalar do operador idêntico. Se $T = cI$, então para dois quais vetores linearmente independentes α_1 e α_2 em V, temos

$$V = Z(\alpha_1; T) \oplus Z(\alpha_2; T)$$

$$p_1 = p_2 = x - c.$$

Para matrizes, esta análise diz que toda 2×2 matriz sobre o corpo F é semelhante sobre F a exatamente uma matriz dos tipos

$$\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix} \quad \begin{bmatrix} 0 & -c_0 \\ 1 & -c_1 \end{bmatrix}$$

Exemplo 3. Seja T o operador linear sobre R^3 que é representado pela matriz

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

em relação à base ordenada canônica. Como já calculamos anteriormente, o polinômio característico de T é $f = (x-1)(x-2)^2$ e o polinômio minimal de T é $p = (x-1)(x-2)$. Assim, sabemos que na decomposição cíclica de T o T-anulador do primeiro vetor α_1 será p. Como estamos operando num espaço tridimensional, só pode existir mais um vetor, α_2. Este vetor deve gerar um subespaço cíclico de dimensão 1, isto é, deve ser um vetor característico de T. O seu T-anulador p_2 deve ser $(x-2)$, porque devemos ter $pp_2 = f$. Notemos que isto nos diz imediatamente que a matriz A é semelhante à matriz

$$B = \begin{bmatrix} 0 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

ou seja, que T é representado por B em relação a uma certa base ordenada. Como podemos encontrar vetores adequados α_1 e α_2? Bem, sabemos que todo vetor que gera um subespaço T-cíclico de dimensão 2 é um α_1 adequado. Portanto, experimentamos ε_1. Temos

$$T\varepsilon_1 = (5, -1, 3)$$

que não é múltiplo escalar de ε_1; logo $Z(\varepsilon_1; T)$ tem dimensão 2. Este espaço consiste dos vetores $a\,\varepsilon_1 + b(T\varepsilon_1)$:

$$\alpha(1, 0, 0) + b(5, -1, 3) = (a + 5b, -b, 3b)$$

ou seja, todos os vetores (x_1, x_2, x_3) que satisfazem $x_3 = -3x_2$. O que desejamos agora é um vetor α_2 tal que $T\alpha_2 = 2\alpha_2$ e $Z(\alpha_2; T)$ seja disjunto de $Z(\varepsilon_1; T)$. Como α_2 deve ser um vetor característico de T, o espaço $Z(\alpha_2; T)$ será simplesmente o espaço unidimensional gerado por α_2 e então o que exigimos é que α_2 não esteja em $Z(\varepsilon_1; T)$. Se $\alpha = (x_1, x_2, x_3)$, pode-se calcular facilmente que $T\alpha = 2\alpha$ se, e somente se, $x_1 = 2x_2 + 2x_3$. Assim, $\alpha_2 = (2, 1, 0)$ satisfaz $T\alpha_2 = 2\alpha_2$ e gera um subespaço T-cíclico disjunto de $Z(\varepsilon_1; T)$. O leitor deverá verificar diretamente que a matriz de T em relação à base ordenada

$$\{(1, 0, 0), (5, -1, 3), (2, 1, 0)\}$$

é a matriz B acima.

Exemplo 4. Suponhamos que T seja um operador linear diagonalizável sobre V. É interessante relacionar uma decomposição cíclica de T com uma base que diagonaliza a matriz de T. Sejam $c_1, \ldots, c_k$ valores característicos distintos de T e seja V_i o espaço dos vetores característicos associado ao valor característico c_i. Então

$$V = V_1 \oplus \ldots \oplus V_k$$

e se $d_i = \dim V_i$, então

$$f = (x - c_1)^{d_1} \ldots (x - c_k)^{d_k}$$

é o polinômio característico de T. Se α é um vetor em V, é fácil relacionar o subespaço cíclico $Z(\alpha; T)$ com os subespaços $V_1, \ldots, V_k$. Existem vetores $\beta_1, \ldots, \beta_k$ determinados de modo único, tais que β_i esteja em V_i e

$$\alpha = \beta_1 + \ldots + \beta_k.$$

Como $T\beta_i = c_i\beta_i$ temos

$$f(T)\alpha = f(c_1)\beta_1 + \ldots + f(c_k)\beta_k \tag{7-20}$$

para todo polinômio f. Dados escalares arbitrários $t_1, \ldots, t_k$ existe um polinômio f tal que $f(c_i) = t_i$, $1 \le i \le k$. Portanto, $Z(\alpha; T)$ é

exatamente o subespaço gerado pelos vetores $\beta_1, \ldots, \beta_k$. Qual é o anulador de α? De acordo com (7-20) temos $f(T)\alpha = 0$ se, e somente, se $f(c_i)\beta_i = 0$ para cada i. Em outras palavras, $f(T)\alpha = 0$ desde que $f(c_i) = 0$ para cada i tal que $\beta_i \neq 0$. Conseqüentemente, o anulador de α é o produto

$$\prod_{\beta_i \neq 0} (x - c_i). \tag{7-21}$$

Seja $\mathscr{B}_i = \{\beta_1^i, \ldots, \beta_{d_i}^i\}$ uma base ordenada de V_i. Seja

$$r = \max_i \, d_i.$$

Definamos os vetores $\alpha_1, \ldots, \alpha_r$ por

$$\alpha_j = \sum_{d_i \geq j} \beta_j^i \qquad 1 \leq j \leq r. \tag{7-22}$$

O subespaço cíclico $Z(\alpha_j; T)$ é o subespaço gerado pelos vetores β_j^i quando i tomar aqueles valores para os quais $d_i \geq j$. O T-anulador de α_i é

$$p_j = \prod_{d_i \geq j} (x - c_i). \tag{7-23}$$

Temos

$$V = Z(\alpha_1; T) \oplus \ldots \oplus Z(\alpha_r; T)$$

pois cada β_j^i pertence a um, e somente um, dos subespaços $Z(\alpha_1; T)$, $\ldots$, $Z(\alpha_1; T)$ e $\mathscr{B} = (\mathscr{B}_1, \ldots, \mathscr{B}_k)$ é uma base de V. Por (7-23), p_{j+1} divide p_j.

Exercícios

1. Seja T o operador linear sobre F^2 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Seja $\alpha_1 = (0, 1)$. Mostrar que $F^2 \neq Z(\alpha_1; T)$ e que não existe nenhum vetor não-nulo α_2 em F^2 tal que $Z(\alpha_2; T)$ seja disjunto de $Z(\alpha_1; T)$.

2. Seja T um operador linear sobre o espaço V de dimensão finita e seja R a imagem de T.

(a) Demonstrar que R possui um subespaço suplementar T-invariante se, e somente se, R é disjunto do núcleo N de T.

(b) Se R e N são disjuntos, demonstrar que N é o único subespaço T-invariante que é um suplementar de R.

3. Seja T o operador linear sobre R^3 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Seja W o núcleo de $T - 2I$. Demonstrar que W não possui nenhum subespaço suplementar T-invariante. [*Sugestão*: Seja $\beta = \varepsilon_1$ e observemos que $(T - 2I)\beta$ está em W. Demonstrar que não existe nenhum α em W tal que $(T - 2I)\beta = (T - 2I)\alpha$.]

4. Seja T o operador linear sobre F^4 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} c & 0 & 0 & 0 \\ 1 & c & 0 & 0 \\ 0 & 1 & c & 0 \\ 0 & 0 & 1 & c \end{bmatrix}.$$

Seja W o núcleo de $T - cI$.

(a) Demonstrar que W é o subespaço gerado por ε_4.

(b) Determinar os geradores unitários dos ideais $S(\varepsilon_4; W)$, $S(\varepsilon_3; W)$, $S(\varepsilon_2; W)$, $S(\varepsilon_1; W)$.

5. Seja T um operador linear sobre o espaço vetorial V sobre o corpo F. Se f é um polinômio sobre F e α está em V, seja $f\alpha = f(T)\alpha$. Suponhamos que $V_1, \ldots, V_k$ sejam subespaços T-invariantes e que $V = V_1 \oplus \ldots \oplus V_k$. Demonstrar que

$$fV = fV_1 \oplus \ldots \oplus fV_k.$$

6. Sejam T, V e F como no Exercício 5. Suponhamos que α e β sejam vetores em V que têm o mesmo T-anulador. Demonstrar que, para todo polinômio f, os vetores $f\alpha$ e $f\beta$ têm o mesmo T-anulador.

7. Determinar os polinômios minimais e as formas racionais de cada uma das seguintes matrizes reais:

$$\begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} c & 0 & -1 \\ 0 & c & 1 \\ -1 & 1 & c \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}$$

8 Seja T o operador linear sobre R^3 que é representado em relação à base ordenada canônica por

$$\begin{bmatrix} 3 & -4 & -4 \\ -1 & 3 & 2 \\ 2 & -4 & -3 \end{bmatrix}.$$

Determinar vetores não-nulos $\alpha_1, \ldots, \alpha_r$ que satisfaçam as condições do Teorema 3.

9. Seja A a matriz real

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 1 & 3 \\ -3 & -3 & -5 \end{bmatrix}.$$

Determinar uma 3×3 matriz real inversível P tal que $P^{-1}AP$ esteja sob a forma racional.

10. Seja F um subcorpo do corpo dos números complexos e seja T o operador linear sobre F^4 que é representado em relação à base ordenada canônica pela matriz

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & a & 2 & 0 \\ 0 & 0 & b & 2 \end{bmatrix}.$$

Determinar o polinômio característico de T. Considerar os casos $a = b = 1$; $a = b = 0$; $a = 0$, $b = 1$. Em cada um destes casos, determinar o polinômio minimal de T e vetores não-nulos $\alpha_1, \ldots, \alpha_r$ que satisfaçam as condições do Teorema 3.

11. Demonstrar que se A e B são 3×3 matrizes sobre o corpo F, uma condição necessária e suficiente para que A e B sejam semelhantes sobre F é que possuam o mesmo polinômio característico e o mesmo polinômio minimal. Dar um exemplo que mostra que isto é falso para 4×4 matrizes.

12. Seja F um subcorpo do corpo dos números complexos e sejam A e B $n \times n$ matrizes sobre F. Demonstrar que se A e B são semelhantes sobre o corpo dos números complexos, então elas são semelhantes sobre F. (*Sugestão*: Demonstrar que a forma racional de A é a mesma seja A considerada como uma matriz sobre F ou como uma matriz sobre C; o mesmo para B.)

13. Seja A uma $n \times n$ matriz com elementos complexos. Demonstrar que se todo valor característico de A é real, então A é semelhante a uma matriz com elementos reais.

14. Seja T um operador linear sobre o espaço V de dimensão finita. Demonstrar que existe um vetor α em V com a seguinte propriedade: Se f é um polinômio e $f(T)\alpha = 0$, então $f(T) = 0$. (Um tal vetor α é denominado um vetor **separador** para a álgebra dos polinômios em T.) Para o caso em que T possui um vetor cíclico, demonstrar diretamente que todo vetor cíclico é um vetor separador para a álgebra dos polinômios em T.

15. Seja F um subcorpo do corpo dos números complexos e seja A uma $n \times n$ matriz sobre F. Seja p o polinômio minimal de A. Se considerarmos A como uma matriz sobre C, então A possuirá um polinômio minimal f, quando considerada como uma $n \times n$ matriz sobre C. Usar um teorema sobre equações lineares para demonstrar que $p = f$. De que forma este resultado decorre do teorema da decomposição cíclica?

16. Seja A uma $n \times n$ matriz com elementos *reais* tal que $A^2 + I = 0$. Demonstrar que n é par e que, se $n = 2k$, então A é semelhante sobre o corpo dos números reais a uma matriz da forma em blocos

$$\begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}$$

onde I é a $k \times k$ matriz unidade.

17. Seja T um operador linear sobre um espaço vetorial V de dimensão finita. Suponhamos que

(a) o polinômio minimal de T seja uma potência de um polinômio irredutível;

(b) o polinômio minimal seja igual ao polinômio característico.

Demonstrar que nenhum subespaço não-trivial e T-invariante possui um subespaço suplementar T-invariante.

18. Se T for um operador linear diagonalizável, então todo subespaço T-invariante terá um subespaço suplementar T-invariante.

19. Seja T um operador linear sobre o espaço V de dimensão finita. Demonstrar que T possui um vetor cíclico se, e somente se, vale o seguinte: Todo operador linear U que comuta com T é um polinômio em T.

20. Seja V um espaço vetorial de dimensão finita sobre F e seja T um operador linear sobre V. Perguntamos quando é que todo vetor não-nulo em V é um vetor cíclico de T. Demonstrar que isto ocorre se, e somente se, o polinômio característico de T é irredutível sobre F.

21. Seja A uma $n \times n$ matriz com elementos *reais*. Seja T o operador linear sobre R^n que é representado por A em relação à base ordenada canônica e seja U o operador linear sobre C^n que é representado por A em relação à base ordenada canônica. Usar o resultado do Exercício 20 para demonstrar o seguinte: Se os únicos subespaços invariantes sob T são R^n e o subespaço nulo, então U é diagonalizável.

7.3 A Forma de Jordan

Suponhamos que N seja um operador linear nilpotente sobre o espaço V de dimensão finita. Consideremos a decomposição cíclica de N que obtemos por meio do Teorema 3. Temos um

inteiro positivo r e r vetores não-nulos $\alpha_1, \ldots, \alpha_r$ em V com N-anuladores $p_1, \ldots, p_r$, tais que

$$V = Z(\alpha_1; N) \oplus \ldots \oplus Z(\alpha_r; N)$$

e p_{i+1} divide p_i para $i = 1, \ldots, r-1$. Como N é nilpotente, o polinômio minimal é x^k para um certo $k \leq n$. Assim, cada p_i é da forma $p_i = x^{k_i}$ e a condição de divisibilidade diz simplesmente que

$$k_1 \geq k_2 \geq \ldots \geq k_r.$$

Evidentemente, $k_r = k$ e $k_r \geq 1$. A matriz associada a x^{k_i} é a $k_i \times k_i$ matriz

(7-24)
$$A_i = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}.$$

Assim, o Teorema 3 nos fornece uma base ordenada de V em relação à qual a matriz de N é a soma direta das matrizes nilpotentes elementares (7-24), cujas dimensões diminuem à medida que i aumenta. Vê-se, a partir disto, que estão associados a uma $n \times n$ matriz nilpotente um inteiro positivo r e r inteiros positivos $k_1, \ldots, k_r$ tais que $k_1 + \ldots + k_r = n$ e $k_i \geq k_{i+1}$, e estes inteiros positivos determinam a forma racional da matriz, isto é, determinam a matriz a menos de semelhança.

Eis aqui algo que gostaríamos de ressaltar sobre o operador nilpotente N acima. O inteiro positivo r é precisamente a nulidade de N; na verdade, uma base do núcleo é formada pelos r vetores

(7-25)
$$N^{k_i - 1}\alpha_i.$$

De fato, suponhamos que α esteja no núcleo de N. Podemos escrever α sob a forma

$$\alpha = f_1\alpha_1 + \ldots + f_r\alpha_r$$

onde f_i é um polinômio, cujo grau podemos supor menor que k_i. Como $N\alpha = 0$, para cada i temos

$$\begin{aligned} 0 &= N(f_i\alpha_i) \\ &= Nf_i(N)\alpha_i \\ &= (xf_i)\alpha_i. \end{aligned}$$

Assim, xf_i é divisível por x^{k_i} e como $\text{gr}(f_i) > k_i$ isto significa que

$$f_i = c_i x^{k_i - 1}$$

onde c_i é um certo escalar. Mas então

$$\alpha = c_1(x^{k_1 - 1}\alpha_1) + \dots + c_r(x^{k_r - 1}\alpha_r)$$

o que nos mostra que os vetores (7-25) formam uma base do núcleo de N. O leitor deverá notar que este fato também é evidente do ponto de vista de matrizes.

O que desejamos fazer agora é combinar nossas conclusões a respeito de operadores ou matrizes nilpotentes com o teorema da decomposição primária do Capítulo 6. A situação é a seguinte: Suponhamos que T seja um operador linear sobre V e que o polinômio característico de T se decomponha sobre F como segue:

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$$

onde $c_1, \dots, c_k$ são elementos distintos em F e $d_i \geq 1$. Então, o polinômio minimal de T será

$$p = (x - c_1)^{r_1} \dots (x - c_k)^{r_k}$$

onde $1 \leq r_i \leq d_i$. Se W_i é o núcleo de $(T - c_iI)^{r_i}$, então o teorema da decomposição primária nos diz que

$$V = W_1 \oplus \dots \oplus W_k$$

e que o operador T_i induzido sobre W_i por T possui polinômio minimal igual a $(x - c_i)^{r_i}$. Seja N_i o operador linear sobre W_i definido por $N_i = T_i - c_iI$. Então, N_i é nilpotente e seu polinômio minimal é x^{r_i}. Sobre W_i, T age como N_i mais o escalar c_i vezes o operador idêntico. Suponhamos que tomemos uma base do subes-

paço W_i correspondente à decomposição cíclica do operador nilpotente N_i. Então a matriz de T_i em relação a esta base ordenada será a soma direta das matrizes

$$\text{(7-26)} \qquad \begin{bmatrix} c & 0 & \dots & 0 & 0 \\ 1 & c & \dots & 0 & 0 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \dots & 1 & c \end{bmatrix}$$

cada uma com $c = c_i$. Além disso, as dimensões destas matrizes diminuem quando se lê da esquerda para a direita. Uma matriz da forma (7-26) é dita uma **matriz elementar de Jordan com valor característico** c. Reunindo todas as bases dos W_i obtemos uma base de V. Descrevamos a matriz A de T em relação a esta base ordenada, a matriz A é a soma direta

$$\text{(7-27)} \qquad A = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & A_k \end{bmatrix}$$

das matrizes $A_1, \dots, A_k$. Cada A_i é da forma

$$A_i = \begin{bmatrix} J_1^{(i)} & 0 & \dots & 0 \\ 0 & J_2^{(i)} & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & J_{n_i}^{(i)} \end{bmatrix}$$

onde cada $J_j^{(i)}$ é uma matriz elementar de Jordan com valor característico c_i. Além disso, dentro de cada A_i, as dimensões das matrizes $J_j^{(i)}$ diminuem à medida que j aumenta. Diremos que uma $n \times n$ matriz A que satisfaz todas as condições descritas até agora neste parágrafo (para certos escalares distintos $c_1, \dots, c_k$) está sob a **forma de Jordan**.

Acabamos de salientar que se T é um operador linear para o qual o polinômio característico se decompõe completamente sobre o corpo de escalares, então existe uma base ordenada de V em relação à qual T é representado por uma matriz que está sob a forma de Jordan. Gostaríamos de mostrar agora que esta matriz é algo associada de modo único a T, a menos da ordem em que os valores característicos de T são escritos. Em outras palavras, se duas matrizes estão sob a forma de Jordan e se elas são semelhantes, então elas podem diferir apenas quanto à ordem dos escalares c_i.

Podemos ver a unicidade como segue. Suponhamos que exista alguma base ordenada de V em relação à qual T seja representado pela matriz de Jordan A descrita no parágrafo anterior. Se A_i é uma $d_i \times d_i$ matriz, então d_i é evidentemente a multiplicidade de c_i como uma raiz do polinômio característico de A, ou de T. Em outras palavras, o polinômio característico de T é

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}.$$

Isto mostra que $c_1, \dots, c_k$ e $d_1, \dots, d_k$ são únicos, a menos da ordem em que são escritos. O fato de que A é a soma direta das matrizes A_i nos fornece uma decomposição em soma direta $V = W_1 \oplus \dots \oplus W_k$ invariante sob T. Observemos agora que W_i deve ser o núcleo de $(T - c_iI)^n$, sendo $n = \dim V$; de fato, $A_i - c_iI$ é obviamente nilpotente e $A_j - c_iI$ é não-singular para $j \neq i$. Portanto, vemos que os subespaços W_i são únicos. Se T_i é o operador induzido sobre W_i por T, então a matriz A_i é determinada de um único modo como a forma racional de $(T_i - c_iI)$.

Desejamos agora fazer mais algumas observações sobre o operador T e a matriz de Jordan A que representa T em relação a uma certa base ordenada. Faremos uma cadeia de observações:

(1) Todo elemento de A que não esteja na diagonal principal ou imediatamente abaixo dela é nulo. Na diagonal de A aparecem os k valores característicos distintos $c_1, \dots, c_k$ de T. Além disso, c_i se repete d_i vezes, sendo d_i a multiplicidade de c_i como uma raiz do polinômio característico, isto é, $d_i = \dim W_i$.

(2) Para cada i, a matriz A_i é a soma direta de n_i matrizes elementares de Jordan, $J_j^{(i)}$, com valor característico c_i. O número n_i é exatamente a dimensão do espaço dos vetores característicos associados ao valor característico c_i. De fato, n_i é o número de blocos nilpotentes elementares na forma racional de $(T_i - c_iI)$

sendo portanto igual à dimensão do núcleo de $(T - c_i I)$. Em particular, notemos que T é diagonalizável se, e somente se, $n_i = d_i$ para todo i.

(3) Para cada i, o primeiro bloco $J^{(i)}$ na matriz A_i é uma $r_i \times r_i$ matriz, sendo r_i a multiplicidade de c_i como uma raiz do polinômio *minimal* de T. Isto decorre do fato de que o polinômio minimal do operador nilpotente $(T_i - c_i I)$ é x^{r_i}.

É claro que temos, como sempre, o mesmo resultado para matrizes. Se B é uma $n \times n$ matriz sobre o corpo F e se o polinômio característico de B se decompõe completamente sobre F, então B é semelhante sobre F a uma $n \times n$ matriz A sob a forma de Jordan e A é única a menos da ordem dos valores característicos. Dizemos que A é a **forma de Jordan** de B.

Além disso, notemos que se F é um corpo algebricamente fechado, então as observações acima se aplicam a todo operador linear sobre um espaço de dimensão finita, ou, a toda $n \times n$ matriz sobre F. Assim, por exemplo, toda $n \times n$ matriz sobre o corpo dos números complexos é semelhante a uma matriz essencialmente única sob a forma de Jordan.

Exemplo 5. Suponhamos que T seja um operador linear sobre C^2. O polinômio característico de T é $(x - c_1)(x - c_2)$, sendo c_1 e c_2 números complexos distintos ou então é $(x - c)^2$. No primeiro caso, T é diagonalizável e é representado em relação a alguma base ordenada por

$$\begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}.$$

No segundo caso, o polinômio minimal de T pode ser $(x - c)$, e então $T = cI$ ou pode ser $(x - c)^2$ e então T é representado em relação a alguma base pela matriz

$$\begin{bmatrix} c & 0 \\ 1 & c \end{bmatrix}.$$

Assim, toda 2×2 matriz sobre o corpo dos números complexos é semelhante a uma matriz de um dos dois tipos acima exibidos, possivelmente com $c_1 = c_2$.

Exemplo 6. Seja A a 3×3 matriz complexa

$$\begin{bmatrix} 2 & 0 & 0 \\ a & 2 & 0 \\ b & c & -1 \end{bmatrix}.$$

O polinômio característico de A é obviamente $(x-2)^2(x+1)$. Ou este polinômio é o polinômio minimal e A é semelhante a

$$\begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

ou então o polinômio minimal é $(x-2)(x+1)$, caso em que A é semelhante a

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Ora,

$$(A-2I)(A+I) = \begin{bmatrix} 0 & 0 & 0 \\ 3a & 0 & 0 \\ ac & 0 & 0 \end{bmatrix}$$

e assim A é semelhante a uma matriz diagonal se, e somente se, $a = 0$.

Exemplo 7. Seja

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & a & 2 \end{bmatrix}.$$

O polinômio característico de A é $(x-2)^4$. Como A é a soma direta de duas 2×2 matrizes, é evidente que o polinômio minimal de A é $(x-2)^2$. Ora, se $a = 0$ ou se $a = 1$, então a matriz A está sob a forma de Jordan. Notemos que as duas matrizes que se obtêm para $a = 0$ e $a = 1$ têm o mesmo polinômio característico e o mesmo polinômio minimal, mas não são semelhantes. Elas não

são semelhantes porque, para a primeira matriz, o espaço-solução de $(A - 2I)$ tem dimensão 3, enquanto que para a segunda matriz a dimensão é 2.

Exemplo 8. As equações diferenciais com coeficientes constantes (Exemplo 14, Capítulo 6) oferecem uma boa ilustração da forma de Jordan. Sejam $a_0, \ldots, a_{n-1}$ números complexos e seja V o espaço de todas as funções f, n vezes diferenciáveis sobre um intervalo da reta real, que satisfaçam a equação diferencial

$$\frac{d^n f}{dx^n} + a_{n-1}\frac{d^{n-1}f}{dx^{n-1}} + \ldots + a_1\frac{df}{dx} + a_0 f = 0.$$

Seja D o operador derivação. Então V é invariante sob D porque V é o núcleo de $p(D)$, onde

$$p = x^n + \ldots + a_1 x + a_0.$$

Qual é a forma de Jordan do operador derivação sobre V?

Sejam $c_1, \ldots, c_k$ as raízes complexas distintas de p:

$$p = (x - c_1)^{r_1} \ldots (x - c_k)^{r_k}.$$

Seja V_i o núcleo de $(D - c_i I)^{r_i}$, isto é, o conjunto das soluções da equação diferencial

$$(D - c_i I)^{r_i} f = 0.$$

Como observamos no Exemplo 15, Capítulo 6, o teorema da decomposição primária nos diz que

$$V = V_1 \oplus \ldots \oplus V_k.$$

Seja N_i a restrição de $D - c_i I$ a V_i. A forma de Jordan do operador D (sobre V) é então determinada pelas formas racionais dos operadores nilpotentes $N_1, \ldots, N_k$ sobre os espaços $V_1, \ldots, V_k$.

Assim, o que devemos conhecer (para vários valores de c) é a forma racional do operador $N = (D - cI)^r$ sobre o espaço V_c, que é formada pelas soluções da equação

$$(D - cI)^r f = 0.$$

Quantos blocos nilpotentes elementares existem na forma racional de N? O número será a nulidade de N, isto é, a dimensão do espaço característico associado ao valor característico c. Essa dimensão é 1 porque qualquer função que satisfaça a equação diferencial

$$Df = cf$$

é um múltiplo escalar da função exponencial $h(x) = e^{cx}$. Portanto, o operador N (sobre o espaço V_c) possui um vetor cíclico. Uma boa escolha para um vetor cíclico é $g = x^{r-1}h$:

$$g(x) = x^{r-1}e^{cx}.$$

Isso nos dá

$$\begin{aligned} Ng &= (r-1)x^{r-2}h \\ &\vdots \\ N^{r-1}g &= (r-1)!\,h \end{aligned}$$

O parágrafo anterior nos mostra que a forma de Jordan de D (sobre o espaço V) é a soma direta de k matrizes elementares de Jordan, uma para cada raiz c_i.

Exercícios

1. Sejam N_1 e N_2 3×3 matrizes nilpotentes sobre o corpo F. Demonstrar que N_1 e N_2 são semelhantes se, e somente se, possuem o mesmo polinômio minimal.

2. Usar o resultado do Exercício 1 e a forma de Jordan para demonstrar o seguinte: Sejam A e B $n \times n$ matrizes sobre o corpo F que possuam o *mesmo* polinômio característico

$$f = (x - c_1)^{d_1}, \ldots, (x - c_k)^{d_k}$$

e o mesmo polinômio minimal. Se nenhum d_i é maior que 3, então A e B são semelhantes.

3. Se A é uma 5×5 matriz complexa com polinômio característico $f = (x-2)^3\,(x+7)^2$ e polinômio minimal $p = (x-2)^2\,(x+7)$, qual é a forma de Jordan de A?

4. Quantas formas de Jordan são possíveis para a 6×6 matriz complexa cujo polinômio característico é $(x+2)^4\,(x-1)^2$?

5. O operador derivação sobre o espaço dos polinômios de grau menor ou igual a 3 é representado em relação à base ordenada "natural" pela matriz

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Qual é a forma de Jordan desta matriz? (F é um subcorpo do corpo dos números complexos.)

6. Seja A a matriz complexa

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 \\ 1 & 1 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Determinar a forma de Jordan de A.

7. Se A é uma $n \times n$ matriz sobre o corpo F com polinômio característico

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k},$$

qual é o traço de A?

8. Classificar, a menos da semelhança, todas as 3×3 matrizes complexas A tais que $A^3 = I$.

9. Classificar, a menos da semelhança, todas as $n \times n$ matrizes complexas A tais que $A^n = I$.

10. Seja n um inteiro positivo, $n \geq 2$ e seja N uma $n \times n$ matriz sobre o corpo F tal que $N^n = 0$ mas $N^{n-1} \neq 0$. Demonstrar que N não possui nenhuma raiz quadrada, isto é, que não existe nenhuma $n \times n$ matriz A tal que $A^2 = N$.

11. Sejam N_1 e N_2 6×6 matrizes nilpotentes sobre o corpo F. Suponhamos que N_1 e N_2 tenham o mesmo polinômio minimal e a mesma nulidade. Demonstrar que N_1 e N_2 são semelhantes. Mostrar que isto não é válido para 7×7 matrizes nilpotentes.

12. Usar o resultado do Exercício 11 e a forma de Jordan para demonstrar o seguintes: Sejam A e B $n \times n$ matrizes sobre o corpo F que possuam o mesmo polinômio característico

$$f = (x - c_1)^{d_1} \dots (x - c_k)^{d_k}$$

e o mesmo polinômio minimal. Suponhamos também que para cada i, os espaços-soluções de $(A - c_i I)$ e de $(B - c_i I)$ tenham a mesma dimensão. Se nenhum dos d_i é maior que 6, então A e B são semelhantes.

13. Se N é uma $k \times k$ matriz nilpotente elementar, isto é, $N^k = 0$ mas $N^{k-1} \neq 0$, mostrar que N^t é semelhante a N. Usar a forma de Jordan para demonstrar que toda $n \times n$ matriz complexa é semelhante à sua transposta.

14. O que está errado na demonstração que segue? Se A é uma $n \times n$ matriz complexa tal que $A^t = -A$, então A é 0. (*Demonstração*: Seja J a forma de Jordan de A. Como $A^t = -A$, $J^t = -J$. Mas J é triangular, logo $J^t = -J$ implica que todo elemento de J é nulo. Como $J = 0$ e A é semelhante a J, vemos que $A = 0$.) (Dar um exemplo de uma A não-nula tal que $A^t = -A$.)

15. Se N é uma 3×3 matriz nilpotente sobre C, demonstrar que $A = I + \frac{1}{2}N - \frac{1}{8}N^2$ satisfaz $A^2 = I + N$, isto é, A é uma raiz quadrada $I + N$. Usar a série binomial $(1 + t)^{1/2}$ para obter uma fórmula semelhante para a raiz quadrada de $I + N$, onde N é uma $n \times n$ matriz nilpotente arbitrária sobre C.

16. Usar o resultado do Exercício 15 paja demonstrar que se c é um número complexo não-nulo e N é uma matriz complexa nilpotente, então $(cI + N)$ possui uma raiz quadrada. Usar depois a forma de Jordan para demonstrar que toda $n \times n$ matriz complexa não-singular possui uma raiz quadrada.

7.4 Cálculo dos Fatores Invariantes

Suponhamos que A seja uma $n \times n$ matriz com elementos no corpo F. Queremos descobrir um processo que permita calcular os fatores invariantes $p_1, \ldots, p_r$ que definem a forma racional de A. Comecemos com o caso muito simples no qual A é a matriz associada (7.2) a um polinômio unitário

$$p = x^n + c_{n-1}x^{n-1} + \ldots + c_1 x + c_0.$$

Vimos na Seção 7.1 que p é o polinômio minimal e característico da matriz associada A. Queremos agora efetuar um cálculo direto, que mostre que p é o polinômio característico de A. Nesse caso,

$$xI = A = \begin{bmatrix} x & 0 & 0 & \ldots & 0 & c_0 \\ -1 & x & 0 & \ldots & 0 & c_1 \\ 0 & -1 & x & \ldots & 0 & c_2 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & x & c_{n-2} \\ 0 & 0 & 0 & \ldots & -1 & x + c_{n-1} \end{bmatrix}$$

Adicionemos x vezes a linha n à linha $(n-1)$. Isso removerá o x na posição $(n-1, n-1)$ e não mudará o determinante. Em seguida, adicionemos x vezes a nova linha $(n-1)$ à linha $(n-2)$. Continuemos sucessivamente, até que todos os x da diagonal principal tenham sido removidos por esse processo. O resultado será a matriz

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & x^n + \ldots + c_1 x + c_0 \\ -1 & 0 & 0 & \cdots & 0 & x^{n-1} + \ldots + c_2 x + c_1 \\ 0 & -1 & 0 & \cdots & 0 & x^{n-2} + \ldots + c_3 x + c_2 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 0 & x^2 + c_{n-1}x + c_{n-2} \\ 0 & 0 & 0 & \cdots & -1 & x + c_{n-1} \end{bmatrix}$$

cujo determinante é igual ao determinante de $xI - A$. O elemento no canto superior, à direita, dessa matriz é o polinômio p. Introduzamos zeros na última coluna, adicionando a ela múltiplos apropriados das outras colunas:

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & p \\ -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 0 \end{bmatrix}$$

Multipliquemos cada uma das primeiras $(n-1)$ colunas por -1 e efetuemos $(n-1)$ trocas de colunas adjacentes para levar a atual coluna n para a primeira posição. O efeito total das $2n-2$ mudanças de sinal, é o de deixar o determinante inalterado. Obteremos a matriz

(7-28)
$$\begin{bmatrix} p & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

É claro, então, que $p = \det(xI - A)$.

Mostraremos que para uma $n \times n$ matriz arbitrária, existe uma seqüência de operações sobre linhas e colunas que transformará

$xI - A$ numa matriz muito parecida com (7-28), na qual os fatores invariantes de A aparecerão na diagonal principal. Deixemos completamente claro que operações iremos usar.

Estaremos trabalhando em $F[x]^{m \times n}$, a coleção das $m \times n$ matrizes cujos elementos são polinômios sobre o corpo F. Se M é uma tal matriz, uma operação elementar sobre as linhas de M é uma das operações seguintes:

1. multiplicação de uma linha de M por um escalar não-nulo de F;

2. substituição da r-ésima linha de M pela linha r mais f vezes a linha s, onde f é um polinômio arbitrário sobre F e $r \neq s$;

3. transposição de duas linhas de M.

A operação inversa de uma operação elementar sobre linhas é uma operação elementar sobre linhas, de mesmo tipo. Devemos notar que não poderíamos fazer tal afirmação se admitíssemos polinômios não constantes em (1). Uma $m \times n$ **matriz elementar,** ou seja, uma matriz elementar em $F[x]^{m \times n}$ é uma matriz que pode ser obtida da $m \times m$ matriz unidade por meio de uma única operação elementar sobre linhas. É claro que toda operação elementar sobre as linhas de M pode ser efetuada, multiplicando M, à esquerda, por uma $m \times m$ matriz elementar conveniente; de fato, se e é a operação, então

$$e(M) = e(I)M.$$

Sejam M e N matrizes em $F[x]^{m \times n}$. Dizemos que N é **linha-equivalente** a M se N pode ser obtida de M por uma seqüência finita de operações elementares sobre linhas:

$$M = M_0 \to M_1 \to \dots \to M_k = N.$$

Evidentemente N é linha-equivalente a M se, e somente se, M é linha-equivalente a N, de modo que poderemos dizer "M e N são linhas-equivalentes". Se N é linha-equivalente a M, então

$$N = PM$$

onde a $m \times m$ matriz P é um produto de matrizes elementares:

$$P = E_1, \dots, E_k.$$

Em particular, P é uma matriz inversível e sua inversa é

$$P^{-1} = E_k^{-1} \dots E_1^{-1}.$$

É óbvio que a inversa de E_j provém da operação elementar inversa sobre linhas.

Tudo isso é exatamente igual ao caso das matrizes com elementos em F e é apenas uma cópia dos resultados elementares do Capítulo 1. Assim, o problema que se apresenta a seguir é o de introduzir uma forma em escada, linha-reduzida, para matrizes cujos elementos são polinômios. Aqui encontramos um novo obstáculo. Como obteremos uma matriz linha-reduzida? Inicialmente devemos localizar o primeiro elemento não-nulo da linha 1 e dividir cada elemento dessa linha por esse elemento não-nulo. Isso não é (necessariamente) possível quando os elementos da matriz são polinômios. Como veremos no próximo teorema, poderemos, em certos casos, contornar essa dificuldade; no entanto, não existe uma forma linha-reduzida inteiramente satisfatória para uma matriz arbitrária de $F[x]^{m \times n}$. Se introduzirmos também operações sobre colunas e estudarmos o tipo de equivalência que resulta do uso de ambos os tipos de operações, poderemos obter para cada matriz uma forma canônica muito útil. A ferramenta básica é o seguinte

Lema. *Seja* M *uma matriz em* $\mathrm{F}[\mathrm{x}]^{\mathrm{m} \times \mathrm{n}}$ *que tenha um elemento não-nulo na sua primeira coluna e seja* p *o máximo divisor comum dos elementos da coluna* 1 *de* M. *Então* M *é linha-equivalente a uma matriz* N, *cuja primeira coluna é*

$$\begin{bmatrix} \mathrm{p} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

Demonstração. Vamos provar algo mais do que enunciamos. Vamos mostrar que existe um algoritmo para determinar N, isto é, uma receita que uma máquina poderia usar para calcular N em um número finito de passagens. Inicialmente, precisamos de alguma notação.

Seja M uma $m \times n$ matriz arbitrária com elementos em $F[x]$ que possui uma primeira coluna não-nula

$$M_1 = \begin{bmatrix} f_1 \\ \cdot \\ \cdot \\ \cdot \\ f_m \end{bmatrix}.$$

Definamos

$$l(M_1) = \min_{f_i \neq 0} \operatorname{gr}(f_i) \tag{7-29}$$
$$p(M_1) = \text{m.d.c.}\,(f_1, \ldots, f_m).$$

Seja j um índice tal que $\operatorname{gr}(f_j) = l(M_1)$. Especificamente, seja j o menor índice i para o qual $\operatorname{gr}(f_i) = l(M_1)$. Tentemos dividir cada f_i por f_j:

$$f_i = f_j g_i + r_i, \qquad r_i = 0 \quad \text{ou} \quad \operatorname{gr}(r_i) < \operatorname{gr}(f_j). \tag{7-30}$$

Para cada i diferente de j, substituamos a linha i de M pela linha i menos g_i vezes a linha j. Multipliquemos a linha j pelo inverso do coeficiente dominante de f_j e em seguida permutemos as linhas j e l. O resultado de todas essas operações será uma matriz M' cuja primeira coluna é

$$M_1' = \begin{bmatrix} \tilde{f}_j \\ r_2 \\ \cdot \\ \cdot \\ \cdot \\ r_{j-1} \\ r_1 \\ r_{j+1} \\ \cdot \\ \cdot \\ \cdot \\ r_m \end{bmatrix}.$$

onde $\tilde{f}_j$ é o polinômio unitário obtido pela normalização de f_j, de modo que o coeficiente dominante seja 1. Demos, assim, um pro-

cesso bem definido para associar a cada matriz M uma matriz M' com as seguintes propriedades:

(a) M' é linha-equivalente a M.

(b) $p(M'_1) = p(M_1)$.

(c) Ou $l(M'_1) < l(M_1)$ ou

$$M'_1 = \begin{bmatrix} p(M_1) \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}.$$

Usando (7-30) e (7-31) é fácil verificar (b) e (c). A propriedade (c) é apenas uma outra maneira de afirmar que existe algum índice i tal que $r_i \neq 0$ e $\text{gr}(r_i) < \text{gr}(f_j)$ ou então $r_i = 0$ para todo i e f_j é, portanto, o máximo divisor comum de $f_1, \ldots, f_m$.

A demonstração do lema é, agora, bastante simples. Comecemos com a matriz M e apliquemos o processo acima para obter M'. A propriedade (c) nos diz que M' funcionará como a matriz N do lema ou $l(M'_1) < l(M_1)$. Nesse último caso, apliquemos o processo a M' para obter a matriz $M^{(2)} = (M')'$. Se $M^{(2)}$ não for um N conveniente, formaremos $M^{(3)} = (M^{(2)})'$, e assim por diante. O fato é que as desigualdades estritas

$$l(M_1) > l(M'_1) > l(M_1^{(2)}) > \ldots$$

não podem continuar por muito tempo. Após não mais que $l(M_1)$ iterações do processo, deveremos chegar a uma matriz $M^{(k)}$ que possui as propriedades que procuramos.

Teorema 6. *Seja* P *uma* m × m *matriz com elementos na álgebra dos polinômios* F[x]. *As seguintes condições são equivalentes:*

(i) P *é inversível.*

(ii) *O determinante de* P *é um polinômio constante não-nulo.*

(iii) P *é linha-equivalente a* m × m *matriz unidade.*

(iv) P *é um produto de matrizes elementares.*

Demonstração. Certamente (i) implica (ii) porque a função determinante é multiplicativa e os únicos polinômios inversíveis de $F[x]$ são os polinômios constantes não-nulos. Na realidade,

no Capítulo 5, usamos a adjunta clássica para mostrar que (i) e (ii) são equivalentes. Nosso argumento, agora, oferece uma demonstração diferente de que (i) é uma conseqüência de (ii). Completaremos o ciclo

$$\begin{array}{ccc} \text{(i)} & \rightarrow & \text{(ii)} \\ \uparrow & & \downarrow \\ \text{(iv)} & \rightarrow & \text{(iii)}. \end{array}$$

A única implicação não óbvia é que (iii) decorre de (ii).

Admitamos (ii) e consideremos a primeira coluna de P. Ela contém certos polinômios $p_1, \ldots, p_m$ e

$$\text{m.d.c.}\,(p_1, \ldots, p_m) = 1$$

porque qualquer divisor comum de $p_1, \ldots, p_m$ necessariamente divide (o escalar) det P. Apliquemos a P o lema anterior para obter a matriz

$$\text{(7-32)} \qquad Q = \begin{bmatrix} 1 & a_2 & \cdots & a_m \\ 0 & & & \\ \cdot & & & \\ \cdot & & B & \\ \cdot & & & \\ 0 & & & \end{bmatrix}$$

que é linha equivalente a P. Uma operação elementar sobre linhas muda o determinante de uma matriz de (no máximo) um fator constante não-nulo. Assim, det Q é um polinômio constante não-nulo. Evidentemente a $(m-1) \times (m-1)$ matriz B em (7-32) e a matriz Q têm o mesmo determinante. Portanto, podemos aplicar o último lema a B. Se continuarmos dessa maneira, em m passagens obteremos uma matriz triangular superior

$$R = \begin{bmatrix} 1 & a_2 & \cdots & a_m \\ 0 & 1 & \cdots & b_m \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

que é linha-equivalente a R. Obviamente R é linha-equivalente à $m \times m$ matriz unidade.

Corolário. *Sejam* M *e* N m × n *matrizes com elementos na álgebra dos polinômios* $F[x]$. *Então* N *é linha-equivalente a* M *se, e somente se,*

$$N = PM,$$

onde P *é uma* m × m *matriz inversível com elementos em* $F[x]$.

Definamos, agora, **operações elementares sobre colunas** e **coluna-equivalência** de modo análogo a operações sobre linhas e linha-equivalência. Não precisaremos de um novo conceito de matriz elementar porque a classe das matrizes que podem ser obtidas por meio de uma operação elementar sobre as colunas da matriz unidade é igual à classe obtida, efetuando-se uma única operação elementar sobre suas linhas.

Definição. *A matriz* N *é equivalente à matriz* M *se for possível passar de* M *para* N *por meio de uma seqüência de operações*

$$M = M_0 \to M_1 \to \dots \to M_k = N$$

cada uma sendo uma operação elementar sobre linhas ou uma operação elementar sobre colunas.

Teorema 7. *Sejam* M *e* N m × n *matrizes com elementos na álgebra dos polinômios* $F[x]$. *Então* N *é equivalente a* M *se, e somente se,*

$$N = PMQ,$$

onde P *é uma matriz inversível em* $F[x]^{m \times m}$ *e* Q *uma matriz inversível em* $F[x]^{n \times n}$.

Teorema 8. *Seja* A *uma* n × n *matriz com elementos no corpo* F *e sejam* $p_1, \dots, p_r$ *os fatores invariantes de* A. *Então a matriz* $xI - A$ *é equivalente a* n × n *matriz diagonal cujos elementos diagonais são* $p_1, \dots, p_r, 1, 1, \dots, 1$.

Demonstração. Existe uma $n \times n$ matriz P, inversível, com elementos em F, tal que PAP^{-1} esteja na forma racional, ou seja, tenha a forma em blocos

$$PAP^{-1} = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & A_r \end{bmatrix}$$

onde A_i é a matriz associada ao polinômio p_i. De acordo com o Teorema 7, a matriz

$$P(xI - A)P^{-1} = xI - PAP^{-1} \tag{7-33}$$

é equivalente a $xI - A$. Ora,

$$xI - PAP^{-1} = \begin{bmatrix} xI - A_1 & 0 & \dots & 0 \\ 0 & xI - A_2 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & xI - A_r \end{bmatrix} \tag{7-34}$$

onde os vários I usados são as matrizes unidades de dimensões apropriadas. No início dessa seção mostramos que $xI - A_i$ é equivalente à matriz

$$\begin{bmatrix} p_i & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Usando (7-33) e (7-34) torna-se claro que $xI - A$ é equivalente a uma matriz diagonal que possui na sua diagonal principal os polinômios p_i e $(n - r)$ elementos iguais a 1. Por meio de uma sucessão de transposições de linhas e colunas, poderemos arranjar os elementos da diagonal na ordem que desejarmos, por exemplo, $p_1, \dots, p_r, 1, \dots, 1$.

O Teorema 8 não nos dá um meio eficiente para calcular os divisores elementares $p_1, \dots, p_r$, porque nossa demonstração depende do teorema da decomposição cíclica. Daremos, agora, um

algoritmo explícito para reduzir uma matriz, cujos elementos são polinômios, à forma diagonal. O Teorema 8 sugere que também podemos fazer com que os elementos sucessivos da diagonal principal dividam um ao outro.

Definição. *Seja* N *uma matriz em* $F[x]^{m \times n}$. *Dizemos que* N *está na* **forma normal** (de Smith) se

(a) *todo elemento que não está na diagonal principal de* N *é nulo*;

(b) *na diagonal principal de* N *aparecem (em ordem) polinômios* $f_1, \ldots, f_l$ *tais que* f_k *divida* f_{k+1}, $1 \leq k \leq l-1$.

Na definição, o número l é $l = \min(m,n)$. Os elementos da diagonal principal são $f_k = N_{kk}$, $k = 1, \ldots, l$.

Teorema 9. *Seja* M *uma* $m \times n$ *matriz com elementos na álgebra dos polinômios* $F[x]$. *Então* M *é equivalente a uma matriz* N *na forma normal.*

Demonstração. Se $M = 0$, não há nada a demonstrar. Se $M \neq 0$ daremos um algoritmo para determinar uma matriz M' que seja equivalente a M e que tenha a forma

$$M' = \begin{bmatrix} f_1 & 0 & \cdots & 0 \\ 0 & & & \\ \cdot & & R & \\ \cdot & & & \\ \cdot & & & \\ 0 & & & \end{bmatrix} \tag{7-35}$$

onde R é uma $(m-1) \times (n-1)$ matriz e f_1 divide cada elemento de R. A demonstração estará então terminada porque podemos aplicar o mesmo processo a R e obter f_2, etc.

Seja $l(M)$ o menor dos graus dos elementos não nulos de M. Procuremos a primeira coluna que tenha um elemento de grau $l(M)$ e permutamos essa coluna com a coluna 1. Chamemos de $M^{(0)}$ a matriz resultante. Descrevamos um processo para determinar uma matriz na forma

$$\begin{bmatrix} g & 0 & \cdots & 0 \\ 0 & & & \\ \cdot & & S & \\ \cdot & & & \\ \cdot & & & \\ 0 & & & \end{bmatrix} \tag{7-36}$$

que seja equivalente à $M^{(0)}$. Comecemos, aplicando à matriz $M^{(0)}$ o processo do lema que precede o Teorema 6, processo que indicaremos por PL6. Resultará uma matriz

$$M^{(1)} = \begin{bmatrix} p & a & \dots & b \\ 0 & c & \dots & d \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & e & \dots & f \end{bmatrix} \tag{7-37}$$

Se os elementos $a, \dots, b$ forem todos nulos, ótimo. Se não forem, usaremos o análogo de PL6 para a primeira linha, um processo que indicaremos por PL6′. O resultado será a matriz

$$M^{(2)} = \begin{bmatrix} q & 0 & \dots & 0 \\ a' & c' & \dots & e' \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ b' & d' & \dots & f' \end{bmatrix} \tag{7-38}$$

onde q é o máximo divisor comum de $p, a, \dots, b$. Ao construirmos $M^{(2)}$ é possível que tenhamos perdido a forma conveniente da coluna 1. Se isso aconteceu poderemos aplicar PL6 novamente. Aqui está o ponto: Em não mais de $l(M)$ passagens

$$M^{(0)} \xrightarrow{\text{PL6}} M^{(1)} \xrightarrow{\text{PL6}'} M^{(2)} \xrightarrow{\text{PL6}} \dots \rightarrow M^{(t)}$$

chegaremos, necessariamente, a uma matriz $M^{(t)}$ que possui a forma (7-36) porque em cada passagem sucessiva temos $l(M^{(k+1)}) < l(M^{(k)})$. Indicaremos o processo que acabamos de definir por P7-36:

$$M^{(0)} \xrightarrow{\text{P7-36}} M^{(t)}.$$

Em (7-36) o polinômio g pode ou não dividir todos os elementos de S. Em caso negativo, determinemos a primeira coluna que tenha um elemento não divisível por g e adicionemos essa coluna à coluna 1. A nova primeira coluna conterá g e um elemento $gh + r$ com $r \neq 0$ e $\text{gr}(r) < \text{gr}(g)$. Apliquemos o processo P7-36 e o resultado será uma outra matriz na forma (7-36) onde o grau do g correspondente diminuiu.

Deverá estar claro agora que após um número finito de passagens, obteremos (7-35), isto é, chegaremos a uma matriz na forma (7-36) onde o grau de g não poderá mais ser reduzido.

Queremos mostrar que a forma normal associada a uma matriz M é única. Dois fatos já vistos indicam por que os polinômios $f_1, \dots, f_l$ do Teorema 9 são determinados de modo único por M. Primeiramente, operações elementares sobre linhas ou colunas não mudam o determinante de uma matriz quadrada a não ser por um fator constante não-nulo. Em segundo lugar, operações elementares sobre linhas ou colunas não mudam o máximo divisor comum dos elementos de uma matriz.

Definição. *Seja* M *uma* $m \times n$ *matriz com elementos em* $F[x]$. *Se* $1 \leq k \leq min(m, n)$, *definimos* $\delta_k(M)$ *como sendo o máximo divisor comum dos determinantes de todas as* $k \times k$ *submatrizes de* M.

Lembremos que uma $k \times k$ submatriz de M é a matriz obtida pela remoção de $m - k$ linhas e $n - k$ colunas de M. Em outras palavras, escolhemos certas k-uplas.

$$I = (i_1, \dots, i_k), \qquad 1 \leq i_1 < \dots < i_k \leq m$$
$$J = (j_1, \dots, j_k), \qquad 1 \leq j_1 < \dots < j_k \leq n$$

e olhamos para a matriz formada, usando aquelas linhas e colunas de M. Estamos interessados nos determinantes

$$D_{I,J}(M) = \det \begin{bmatrix} M_{i_1 j_1} & \dots & M_{i_1 j_k} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ M_{i_k j_1} & \dots & M_{i_k j_k} \end{bmatrix}. \tag{7-39}$$

O polinômio $\delta_k(M)$ é o máximo divisor comum dos polinômios $D_{I,J}(M)$ quando I e J percorrem todas as possíveis k-uplas.

Teorema 10. *Se* M *e* N *são* $m \times n$ *matrizes equivalentes com elementos em* $F[x]$, *então*

$$\delta_k(M) = \delta_k(N), \qquad 1 \leq k \leq min(m, n). \tag{7-40}$$

Demonstração. Será suficiente mostrar que uma única operação elementar e sobre linhas não altera δ_k. Como a inversa de e

é também uma operação elementar sobre linhas, bastará mostrar o seguinte. Se um polinômio f divide todo $D_{I,J}(M)$, então f divide $D_{I,J}(e(M))$ para todas as k-uplas I e J.

Como estamos considerando operações sobre linhas, sejam $\alpha_1, \ldots, \alpha_m$ as linhas de M e usemos a notação

$$D_J(\alpha_{i_1}, \ldots, \alpha_{i_k}) = D_{I,J}(M).$$

Dados I e J, qual é a relação entre $D_{I,J}(M)$ e $D_{I,J}(e(M))$? Consideremos os três tipos de operações e:

(a) multiplicação da linha r pelo escalar não-nulo c;

(b) substituição da linha r pela linha r mais g vezes a linha s, $r \neq s$;

(c) transposição das linhas r e s, $r \neq s$.

Esqueçamos por um momento as operações do tipo (c) e concentremo-nos nos tipos (a) e (b) que somente alteram a linha r. Se r não for um dos índices $i_1, \ldots, i_k$, então

$$D_{I,J}(e(M)) = D_{I,J}(M).$$

Se r estiver entre os índices $i_1, \ldots, i_k$, então teremos, nos dois casos,

(a)
$$\begin{aligned} D_{I,J}(e(M)) &= D_J(\alpha_{i_1}, \ldots, c\alpha_r, \ldots, \alpha_{i_k}) \\ &= cD_J(\alpha_{i_1}, \ldots, \alpha_r, \ldots, \alpha_{i_k}) \\ &= cD_{I,J}(M); \end{aligned}$$

(b)
$$\begin{aligned} D_{I,J}(e(M)) &= D_J(\alpha_{i_1}, \ldots, \alpha_r + g\alpha_s, \ldots, \alpha_{i_k}) \\ &= D_{I,J}(M) + gD_J(\alpha_{i_1}, \ldots, \alpha_s, \ldots, \alpha_{i_k}). \end{aligned}$$

Para as operações do tipo (a) é claro que qualquer f que divida $D_{I,J}(M)$ também dividirá $D_{I,J}(e(M))$. Para o caso das operações do tipo (b) observemos que

$$D_J(\alpha_{i_1}, \ldots, \alpha_s, \ldots, \alpha_{i_k}) = 0, \qquad \text{se } s = i_j \text{ para algum } j$$
$$D_J(\alpha_{i_1}, \ldots, \alpha_s, \ldots, \alpha_{i_k}) = \pm D'_{I,J}(M), \qquad \text{se } s \neq i_j \text{ para todo } j.$$

O I' na última equação é a k-upla $(i_1, \ldots, s, \ldots, i_k)$ arranjada em ordem crescente. Deverá ser evidente agora que se f dividir todo $D_{I,J}(M)$ então f dividirá todo $D_{I,J}(e(M))$.

Poderemos tratar o caso das operações do tipo (c) usando praticamente o mesmo argumento ou então usando o fato de que estas operações podem ser efetuadas por meio de uma seqüência de operações do tipo (a) e (b).

Corolário. *Cada matriz* M *em* $F[x]^{m\times n}$ *é equivalente a exatamente uma matriz* N *na forma normal. Os polinômios* $f_1, \ldots, f_l$ *que ocorrem na diagonal principal de* N *são*

$$f_k = \frac{\delta_k(M)}{\delta_{k-1}(M)}, \qquad 1 \leq k \leq min(m, n)$$

onde, por conveniência, definimos $\delta_0(M) = 1$.

Demonstração. Se N está na forma normal com elementos diagonais $f_1, \ldots, f_l$ é fácil ver que

$$\delta_k(N) = f_1 f_2 \ldots f_k.$$

A matriz N do último corolário é denominada **forma normal** de M. Os polinômios $f_1, \ldots, f_l$ são denominados **fatores invariantes** de M.

Suponhamos que A seja uma $n \times n$ matriz com elementos em F e sejam $p_1, \ldots, p_r$ os fatores invariantes de A. Vemos agora que a forma normal da matriz $xI - A$ possui elementos diagonais $1, 1, \ldots, 1, p_r, \ldots, p_1$. O último corolário nos diz o que são $p_1, \ldots, p_r$, em termos das submatrizes de $xI - A$. O número $n - r$ é o maior k tal que $\delta_k(xI - a) = 1$. O polinômio minimal p_1 é o polinômio característico de A, dividido pelo máximo divisor comum dos determinantes de todas as $(n-1) \times (n-1)$ submatrizes de $xI - A$, etc.

Exercícios

1. Verdadeiro ou falso? Toda matriz em $F[x]^{n\times n}$ é linha-equivalente a uma matriz triangular superior.

2. Seja T um operador linear sobre um espaço vetorial de dimensão finita e seja A a matriz de T em relação a alguma base ordenada. Então T possui um vetor cíclico se, e somente se, os determinantes das $(n-1) \times (n-1)$ submatrizes de $xI - A$ são relativamente primos.

3. Seja A uma $n \times n$ matriz com elementos no corpo F e sejam $f_1, \ldots, f_n$ os elementos diagonais da forma normal de $xI - A$. Para que matrizes A tem-se $f_1 \neq 1$?

4 Construir um operador linear T com polinômio minimal $x^2(x-1)^2$ e polinômio característico $x^3(x-1)^4$. Descrever a decomposição primária do espaço vetorial sob T e determinar as projeções sobre as componentes primárias. Determinar uma base em relação à qual a matriz de T esteja sob a forma de Jordan. Determinar também, explicitamente, uma decomposição do espaço em soma direta de subespaços T-cíclicos, como no Teorema 3, e dar os fatores invariantes.

5. Seja T um operador linear sobre R^8, representado em relação à base canônica pela matriz

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & -1 & -1 & -1 & -1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(a) Determinar o polinômio característico e os fatores invariantes.

(b) Determinar a decomposição primária de R^8 sob T e as projeções sobre as componentes primárias. Determinar as decomposições cíclicas de cada componente primária, como no Teorema 3.

(c) Determinar a forma de Jordan de A.

(d) Determinar uma decomposição de R^8 em soma direta de subespaços T-cíclicos, como no Teorema 3. (*Sugestão*: Uma das maneiras de se fazer isso é usar os resultados de (b) e uma generalização apropriada das idéias discutidas no Exemplo 4.)

7.5 Resumo; Operadores Semi-Simples

Nos dois últimos capítulos, estivemos tratando de um único operador linear T sobre um espaço vetorial V de dimensão finita. O programa foi decompor T numa soma direta de operadores lineares de natureza elementar, com o objetivo de obter informações detalhadas sobre como T 'opera' sobre o espaço V. Recordemos rapidamente onde nos encontramos.

Começamos estudando T por meio de valores característicos e vetores característicos. Introduzimos os operadores diagonalizáveis, operadores que podem ser descritos completamente em termos de valores e vetores característicos. Observamos então que podia ocorrer que T não tivesse nenhum vetor característico. Mesmo no caso de um corpo de escalares algebricamente fechado, em que todo operador linear realmente possui pelo menos um vetor característico, notamos que os vetores característicos de T nem sempre geravam o espaço.

Demonstramos então o teorema da decomposição cíclica, exprimindo um operador linear arbitrário como a soma direta de operadores que tinham um vetor cíclico, sem fazer nenhuma hipótese quanto ao corpo de escalares. Se U é um operador linear que tem um vetor cíclico, existe uma base $\{\alpha_1, \ldots, \alpha_n\}$ tal que

$$U\alpha_j = \alpha_{j+1}, \quad j = 1, \ldots, n-1$$
$$U\alpha_n = -c_0\alpha_1 - c_1\alpha_2 - \ldots - c_{n-1}\alpha_n.$$

A ação de U sobre esta base é então a de transformar cada α_j no vetor seguinte α_{j+1}, com a exceção de $U\alpha_n$ que é uma combinação linear predeterminada dos vetores da base. Como um operador linear genérico T é a soma direta de um número finito de tais operadores U, obtivemos uma descrição explícita e razoavelmente elementar da ação de T.

Aplicamos a seguir o teorema da decomposição cíclica a operadores nilpotentes. Para o caso de um corpo de escalares algebricamente fechado, combinamos este resultado com o teorema da decomposição primária obtendo a forma de Jordan. A forma de Jordan fornece uma base $\{\alpha_1, \ldots, \alpha_n\}$ do espaço V tal que, para cada j, ou $T\alpha_j$ é um múltiplo escalar de α_j ou $T\alpha_j = c\alpha_j + \alpha_{j+1}$. Esta base certamente descreve a ação de T de uma maneira explícita e elementar.

A importância da forma racional (ou da forma de Jordan) origina-se do fato de ela existir e não do fato de poder ser determinada em casos particulares. É claro que se se tem um particular operador linear T e se pode determinar a sua forma racional ou de Jordan, deve-se fazê-lo, pois, tendo esta forma, pode-se conseguir vastas quantidades de informações sobre T. Dois tipos de dificuldades surgem no cálculo dessas formas canônicas. Uma dificuldade é, obviamente, a extensão dos cálculos. A outra dificuldade é que pode não existir nenhum método de efetuar os cálculos mesmo que se tenham paciência e tempo suficientes. A segunda dificuldade surge ao, digamos, se tentar determinar a forma de Jordan de uma matriz complexa. Simplesmente não existe nenhum método bem definido de se decompor o polinômio característico e assim já se é barrado no início. A forma racional não apresenta esta dificuldade. Como mostramos na Seção 7.4, existe um método bem definido para se determinar a forma racional de uma dada $n \times n$ matriz; contudo, tais cálculos são usualmente longos demais.

Em nosso resumo dos resultados destes dois últimos capítulos, ainda não mencionamos um teorema que demonstramos. É o teo-

rema que afirma que se T é um operador linear sobre um espaço vetorial de dimensão finita sobre um corpo algebricamente fechado, então T pode ser expresso de um único modo como a soma de um operador diagonalizável com um operador nilpotente os quais comutam. Ele foi demonstrado a partir do teorema da decomposição primária e algumas informações sobre operadores diagonalizáveis. Não é um teorema tão profundo como o teorema da decomposição cíclica ou a existência da forma de Jordan, mas possui aplicações importantes e úteis em certas partes da matemática. Concluindo este capítulo, vamos demonstrar um teorema análogo, sem supor que o corpo de escalares seja algebricamente fechado. Começamos por definir os operadores que desempenharão o papel dos operadores diagonalizáveis.

Definição. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F *e seja* T *um operador linear sobre* V. *Dizemos que* T *é* **semi-simples** *se todo subespaço* T-*invariante possui um subespaço suplementar* T-*invariante.*

O que estamos prestes a demonstrar é que, com certas restrições sobre o corpo F, todo operador linear T pode ser expresso de um único modo como $T = S + N$, sendo S semi-simples, N nilpotente e $SN = NS$. Primeiro, vamos caracterizar os operadores semi-simples por meio de seus polinômios minimais e esta caracterização nos mostrará que, quando F é algebricamente fechado, um operador é semi-simples se, e somente se, é diagonalizável.

Lema. *Seja* T *um operador linear sobre o espaço vetorial* V *de dimensão finita e seja* $V = W_1 \oplus \dots \oplus W_k$ *a decomposição primária de* T. *Em outras palavras, se* p *é o polinômio de* T *e* $p = p_1^{r_1} \dots p_k^{r_k}$ *é a decomposição de* p *em fatores primos, então* W_j *é o núcleo de* $p_j(T)^{r_j}$. *Seja* W *um subespaço arbitrário de* V *que seja invariante sob* T. *Então*

$$W = (W \cap W_1) \oplus \dots \oplus (W \cap W_k).$$

Demonstração. Para a demonstração precisaremos recordar um corolário de nossa demonstração do teorema da decomposição primária na Seção 68. Se $E_1, \dots, E_k$ são as projeções associadas à decomposição $V = W_1 \oplus \dots \oplus W_k$, então cada E_j é um polinômio em T. Isto é, existem polinômios $h_1, \dots, h_k$ tais que $E_j = h_j(T)$.

Seja agora W um subespaço invariante sob T. Se α é um vetor qualquer em W, então $\alpha = \alpha_1 + \dots + \alpha_k$, com α_j em W_j. Ora,

$\alpha_j = E_j\alpha = h_j(T)\alpha$, e como W é invariante sob T, cada α_j também está em W. Assim, cada vetor α em W é da forma $\alpha = \alpha_1 + \ldots + \alpha_k$, com α_j na interseção $W \cap W_j$. Esta expressão é única pois $V = W_1 \oplus \oplus \ldots \oplus W_k$. Portanto,

$$W = (W \cap W_1) \oplus \ldots \oplus (W \cap W_k).$$

Lema. *Seja* T *um operador linear sobre* V *e suponhamos que o polinômio minimal de* T *seja irredutível sobre o corpo* F *de escalares. Então* T *é semi-simples.*

Demonstração. Seja W um subespaço de V que seja invariante sob T. Precisamos demonstrar que W possui um subespaço suplementar T-invariante. De acordo com um corolário do Teorema 3, será suficiente demonstrar que se f é um polinômio e β é um vetor em V tais que $f(T)\beta$ esteja em W, então existe um vetor α em W tal que $f(T)\beta = f(T)\alpha$. Portanto, suponhamos que β esteja em V e que f seja um polinômio tal que $f(T)\beta$ esteja em W. Se $f(T)\beta = 0$, fazemos $\alpha = 0$ e então α é um vetor em W tal que $f(T)\beta = f(T)\alpha$. Se $f(T)\beta \neq 0$, o polinômio f não é divisível pelo polinômio minimal p do operador T. Como p é primo, isto significa que f e p são relativamente primos e existem polinômios g e h tais que $fg + ph = 1$. Como $p(T) = 0$ temos $f(T)g(T) = I$. Daqui segue que o vetor β deve estar no subespaço W; de fato,

$$\begin{aligned}\beta &= g(T)f(T)\beta \\ &= g(T)(f(T)\beta)\end{aligned}$$

enquanto $f(T)\beta$ está em W e W é invariante sob T. Basta tomar $\alpha = \beta$.

Teorema 11. *Seja* T *um operador linear sobre o espaço vetorial* V *de dimensão finita. Uma condição necessária e suficiente para que* T *seja semi-simples é que o polinômio minimal* p *de* T *seja da forma* $p = p_1, \ldots, p_k$, *sendo* $p_1, \ldots, p_k$ *polinômios irredutíveis distintos sobre o corpo* F *de escalares.*

Demonstração. Suponhamos que T seja semi-simples. Mostraremos que nenhum polinômio irredutível se repete na decomposição do polinômio minimal p em fatores primos. Suponhamos o contrário. Então existe um polinômio unitário não-constante g tal que g^2 divide p. Seja W o núcleo do operador $g(T)$. Então W é invariante sob T. Ora, $p = g^2h$ para algum polinômio h. Como g

não é um polinômio constante, o operador $g(T)h(T)$ não é o operador nulo e existe um vetor β em V tal que $g(T)h(T)\beta = 0$, isto é, $(gh)\beta \neq 0$. Ora, $(gh)\beta$ está no subespaço W, pois $g(gh\beta) = g^2h\beta = p\beta = 0$. Mas não existe nenhum vetor α em W tal que $gh\beta = gh\alpha$; de fato, se α está em W

$$(gh)\alpha = (hg)\alpha = h(g\alpha) = h(0) = 0.$$

Assim, W não pode ter um subespaço suplementar T-invariante, contradizendo a hipótese de T ser semi-simples.

Suponhamos agora que a decomposição de p em fatores primos seja $p = p_1, \ldots, p_k$, sendo $p_1, \ldots, p_k$ polinômios unitários (não-constantes), irredutíveis e distintos. Seja W um subespaço de V que seja invariante sob T. Vamos demonstrar que W possui um subespaço suplementar T-invariante. Seja $V = W_1 \oplus \ldots \oplus W_k$ a decomposição primária de T, isto é, seja W_j o núcleo de $p_j(T)$. Seja T_j o operador linear induzido sobre W_j por T, de modo que o polinômio minimal de T_j é o primo p_j. Ora, $W \cap W_j$ é um subespaço de W_j que é invariante sob T_j (ou sob T). Pelo último lema, existe um subespaço V_j de W_j tal que $W_j = (W \cap W_j) \oplus V_j$ e V_j seja invariante sob T_j (e, portanto, sob T). Então temos

$$\begin{aligned} V &= W_1 \oplus \ldots \oplus W_k, \\ &= (W \cap W_1) \oplus V_1 \oplus \ldots \oplus (W \cap W_k) \oplus V_k \\ &= (W \cap W_1) + \ldots + (W \cap W_k) \oplus V_1 \oplus \ldots \oplus V_k. \end{aligned}$$

Pelo primeiro lema acima, $W = (W \cap W_1) \oplus \ldots \oplus (W \cap W_k)$, de modo que se $W' = V_1 \oplus \ldots \oplus V_k$, então $V = W \oplus W'$ e W' é invariante sob T.

Corolário. *Se* T *é um operador linear sobre um espaço vetorial de dimensão finita sobre um corpo algebricamente fechado, então* T *é semi-simples se, e somente se,* T *é diagonalizável.*

Demonstração. Se o corpo F de escalares é algebricamente fechado, os primos unitários sobre F são os polinômios $x - c$. Neste caso, T é semi-simples se, e somente se, o polinômio minimal de T é $p = (x - c_1)\ldots(x - c_k)$, sendo $c_1, \ldots, c_k$ elementos distintos de F. Este é exatamente o critério para a diagonalização de T, por nós estabelecido no Capítulo 6.

Gostaríamos de destacar que T é semi-simples se, e somente se, existe um polinômio f, que seja um produto de primos distintos,

tal que $f(T) = 0$. Isto difere apenas superficialmente da condição de que o polinômio minimal seja um produto de primos distintos.

Voltemos agora ao problema de exprimir um operador linear como a soma de um operador semi-simples e um operador nilpotente que comutem. Para esta parte, restringiremos o corpo de escalares a um subcorpo do corpo dos números complexos. O leitor informado verá que o importante é o corpo F ser um corpo de característica zero, isto é, para cada inteiro positivo n, a soma $1 + \ldots + 1$ (n vezes) em F não deve ser nula. Para um polinômio f sobre F, indiquemos por $f^{(k)}$ a k-ésima derivada formal de f. Em outras palavras, $f^{(k)} = D^k f$ onde D é o operador derivação sobre o espaço dos polinômios. Se g é um outro polinômio, $f(g)$ indica o resultado de se substituir g em f, isto é, o polinômio obtido aplicando f ao elemento g na álgebra linear $F[x]$.

Lema (Fórmula de Taylor). *Seja* F *um corpo de característica zero e sejam* g *e* h *polinômios sobre* F. *Se* f *é um polinômio qualquer sobre* F *com* $gr(f) \leq n$, *então*

$$f(g) = f(h) + f^{(1)}(h)(g - h) + \frac{f^{(2)}(h)}{2!}(g - h)^2 + \ldots + \frac{f^{(n)}(h)}{n!}(g - h)^n.$$

Demonstração. O que estamos demonstrando é uma fórmula de Taylor generalizada. O leitor provavelmente está acostumado a ver o caso particular em que $h = c$, um polinômio constante, e $g = x$. Nesse caso, a fórmula diz:

$$f = f(x) = f(c) + f^{(1)}(c)(x - c) + \\ + \frac{f^{(2)}(c)}{2!}(x - c)^2 + \ldots + \frac{f^{(n)}(c)}{n!}(x - c)^n.$$

A demonstração desta fórmula é simplesmente uma aplicação do teorema binomial

$$(a + b)^k = a^k + ka^{k-1}b + \frac{k(k-1)}{2!}a^{k-2}b^2 + \ldots + b^k.$$

De fato, o leitor notará que, sendo a substituição e a derivação processos lineares, basta demonstrar a fórmula para $f = x^k$. A fórmula para $f = \sum_{k=0}^{n} c_k x^k$ decorre por uma combinação linear. No

caso $f = x^k$ com $k \leq n$, a fórmula diz

$$g^k = h^k + kh^{k-1}(g-h) + \frac{k(k-1)}{2!}h^{k-2}(g-h)^2 + \ldots + (g-h)^k$$

que é exatamente o desenvolvimento binomial de

$$g^k = [h + (g-h)]^k.$$

Lema. *Seja* F *um subcorpo do corpo dos números complexos, seja* f *um polinômio sobre* F *e seja* f′ *a derivada de* f. *As seguintes afirmações são equivalentes*:

(a) f *é um produto de polinômios irredutíveis e distintos sobre* F.

(b) f *e* f′ *são relativamente primos.*

(c) *Considerado como um polinômio com coeficientes complexos,* f *não possui raízes múltiplas.*

Demonstração. Demonstraremos primeiro que (a) e (b) são afirmações equivalentes sobre f. Suponhamos, na decomposição de f em fatores primos sobre o corpo F, que algum polinômio (não-constante) primo p se repita. Então $f = p^2h$ para algum h em $F[x]$. Então,

$$f' = p^2h' + 2pp'h$$

e p é também um divisor de f'. Logo, f e f' não são relativamente primos. Concluímos que (b) implica (a).

Suponhamos agora que $f = p_1, \ldots, p_k$, onde $p_1, \ldots, p_k$ são polinômios não-constantes, irredutíveis e distintos sobre F. Seja $f_j = f/p_j$. Então,

$$f' = p_1'f_1 + p_2'f_2 + \ldots + p_k'f_k.$$

Seja p um polinômio primo que divida f e f'. Então $p = p_i$ para algum i. Ora, p_i divide f_j para $j \neq i$ e como p_i também divide

$$f' = \sum_{j=1}^{k} p_j'f_j$$

vemos que p_i deve dividir $p_i'f_i$. Portanto, p_i divide f_i ou p_i'. Mas p_i não divide f_i uma vez que $p_1, \ldots, p_k$ são distintos. Então, p_i divide p_i'. Isto não é possível, pois o grau de p_i' é um a menos que

o grau de p_i. Concluímos que nenhum primo divide f e f', ou seja, que $(f, f') = 1$.

Para ver que a afirmação (c) é equivalente a (a) e (b), precisamos observar apenas o seguinte. Suponhamos que f e g sejam polinômios sobre F, um subcorpo do corpo dos números complexos. Podemos considerar f e g também como polinômios com coeficientes complexos. A afirmação de que f e g são relativamente primos como polinômios sobre F é equivalente à afirmação de que f e g são relativamente primos como polinômios sobre o corpo dos números complexos. Deixamos a demonstração deste resultado como exercício. Usemos este fato com $g = f'$. Notemos que (c) é exatamente (a) quando f é considerado como um polinômio sobre o corpo dos números complexos. Assim, (b) e (c) são equivalentes, pelo mesmo argumento utilizado acima.

Podemos agora demonstrar um teorema que tornará mais evidente a relação entre operadores semi-simples e operadores diagonalizáveis.

Teorema 12. *Seja* F *um subcorpo do corpo dos números complexos, seja* V *um espaço vetorial de dimensão finita sobre* F *e seja* T *um operador linear sobre* V. *Seja* $\mathscr{B}$ *uma base ordenada de* V *e seja* A *a matriz de* T *em relação à base ordenada* $\mathscr{B}$. *Então,* T *é semi-simples se, e somente se, a matriz* A *é semelhante, sobre o corpo dos números complexos, a uma matriz diagonal.*

Demonstração. Seja p o polinômio minimal de T. De acordo com o teorema 11, T é semi-simples se, e somente se, $p = p_1 \dots p_k$ onde $p_1, \dots, p_k$ são polinômios distintos irredutíveis sobre F. Pelo último lema, temos que T é semi-simples se, e somente se, p não possui raízes complexas múltiplas.

Ora, p também é o polinômio minimal da matriz A. Sabemos que A é semelhante sobre o corpo dos números complexos a uma matriz diagonal se, e somente se, o seu polinômio minimal não possui raízes complexas múltiplas. Isto demonstra o teorema.

Teorema 13. *Seja* F *um subcorpo do corpo dos números complexos, seja* V *um espaço vetorial de dimensão finita sobre* F *e seja* T *um operador linear sobre* V. *Existe um operador semi-simples* S *sobre* V *e um operador nilpotente* N *sobre* V *tais que*

(i) $T = S + N$;

(ii) $SN = NS$.

Além disso, o S *semi-simples e o* N *nilpotente que satisfazem* (i) e (ii) *são únicos e cada um é um polinômio em* T.

Demonstração. Seja $p_1^{r_1} \dots p_k^{r_k}$ a decomposição em fatores primos do polinômio minimal de T e seja $f = p_1 \dots p_k$. Seja r o maior dos inteiros positivos $r_1, \dots, r_k$. Então, o polinômio f é um produto de primos distintos, f^r é divisível pelo polinômio minimal de T e então

$$f(T)^r = 0.$$

Vamos construir uma seqüência de polinômios: $g_0, g_1, g_2, \dots$ tais que

$$f\left(x - \sum_{j=0}^{n} g_j f^j\right)$$

seja divisível por f^{n+1}, $n = 0, 1, 2, \dots$ Tomando $g_0 = 0$ temos que $f(x - g_0 f^0) = f(x) = f$ é divisível por f. Suponhamos que tenhamos escolhido $g_0, \dots, g_{n-1}$. Seja

$$h = x - \sum_{j=0}^{n-1} g_j f^j$$

de modo que, por hipótese, $f(h)$ é divisível por f^n. Queremos tomar g_n de modo que

$$f(h - g_n f^n)$$

seja divisível por f^{n+1}. Aplicando a fórmula geral de Taylor, obtemos

$$f(h - g_n f^n) = f(h) - g_n f^n f'(h) + f^{n+1} b$$

onde b é algum polinômio. Por hipótese, $f(h) = qf^n$. Assim, vemos que para $f(h - g_n f^n)$ ser divisível por f^{n+1} basta escolher g_n de maneira tal que $(q - g_n f')$ seja divisível por f. Isto pode ser feito, pois f não possui fatores primos repetidos e então f e f' são relativamente primos. Se a e e são polinômios tais que $af + ef' = 1$ e se fizermos $g_n = eq$, então $q - g_n f'$ será divisível por f.

Agora temos uma seqüência $g_0, g_1, \ldots$ tal que f^{n+1} divide $f\left(x - \sum_{j=0}^{n} g_j f^j\right)$. Tomemos $n = r - 1$; como $f(T)^r = 0$, temos

$$f\left(T - \sum_{j=0}^{r-1} g_j(T) f(T)^j\right) = 0.$$

Seja

$$N = \sum_{j=1}^{r-1} g_j(T) f(T)^j = \sum_{j=0}^{r-1} g_j(T) f(T)^j.$$

Como $\sum_{j=1}^{n} g_j f^j$ é divisível por f, vemos que $N^r = 0$ e N é nilpotente. Seja $S = T - N$. Então $f(S) = f(T - N) = 0$. Como f possui fatores primos distintos, S é semi-simples.

Temos agora $T = S + N$ onde S é semi-simples, N é nilpotente e cada um é um polinômio em T. Para demonstrar a afirmação da unicidade, passaremos do corpo de escalares F ao corpo dos números complexos. Seja $\mathscr{B}$ uma base ordenada do espaço V. Então temos

$$[T]_{\mathscr{B}} = [S]_{\mathscr{B}} + [N]_{\mathscr{B}}$$

sendo $[S]_{\mathscr{B}}$ diagonalizável sobre o corpo dos números complexos e $[N]_{\mathscr{B}}$ nilpotente. Esta matriz diagonalizável e esta matriz nilpotente que comutam são determinadas de modo único, como demonstramos no Capítulo 6.

Exercícios

1. Se N é um operador linear nilpotente sobre V, mostrar que para todo polinômio f, a parte semi-simples de $f(N)$ é um múltiplo escalar do operador idêntico (F é um subcorpo de C).

2. Seja F um subcorpo do corpo dos números complexos, V um espaço vetorial de dimensão finita sobre F e T um operador linear semi-simples sobre V. Se f é um polinômio arbitrário sobre F, demonstrar que $f(T)$ é semi-simples.

3. Seja T um operador linear sobre um espaço de dimensão finita sobre um subcorpo de C. Demonstrar que T é semi-simples se, e somente se, vale o seguinte: Se f é um polinômio e $f(T)$ é nilpotente, então $f(T) = 0$.

Capítulo 8
ESPAÇOS COM PRODUTO INTERNO

8.1 Produtos Internos

Em todo este capítulo trataremos apenas de espaços vetoriais reais ou complexos, isto é, de espaços vetoriais sobre o corpo dos números reais ou sobre o corpo dos números complexos. Nosso objetivo principal é estudar espaços vetoriais nos quais tenha sentido falar do "comprimento" de um vetor e do "ângulo" entre dois vetores. Faremos isto por meio do estudo de um certo tipo de função definida sobre pares de vetores e tomando valores escalares, conhecida como um produto interno. Um exemplo de produto interno é o produto escalar de vetores em R^3. O produto escalar dos vetores

$$\alpha = (x_1, x_2, x_3) \quad \text{e} \quad \beta = (y_1, y_2, y_3)$$

em R^3 é o número real

$$(\alpha\beta) = x_1y_1 + x_2y_2 + x_3y_3.$$

Geometricamente, este produto escalar é o produto do comprimento de α pelo comprimento de β e pelo cosseno do ângulo entre α e β. Assim, é possível definir os conceitos geométricos de "comprimento" e "ângulo" em R^3 em termos do produto escalar que é algebricamente definido.

Um produto interno sobre um espaço vetorial é uma função com propriedades semelhantes às do produto escalar em R^3 e, em termos de tal produto interno, pode-se também definir "comprimento" e "ângulo". Nossos comentários sobre a noção geral de ângulo restringir-se-ão ao conceito de perpendicularidade (ou ortogonalidade) de vetores. Nesta primeira seção, vamos dizer o que é um produto interno, considerar alguns exemplos particulares e estabelecer algumas propriedades básicas dos produtos internos.

Então, voltar-nos-emos ao trabalho de discutir comprimento e ortogonalidade.

Definição. *Seja* F *o corpo dos números reais ou o corpo dos números complexos e seja* V *um espaço vetorial sobre* F. *Um* **produto interno** *sobre* V *é uma função que associa a cada par ordenado de vetores* α, β *em* V *um escalar* $(\alpha|\beta)$ *em* F *de maneira tal que para todo* α, β, γ *em* V *e todos escalares* c

(a) $(\alpha + \beta|\gamma) = (\alpha|\gamma) + (\beta|\gamma)$;

(b) $(c\alpha|\beta) = c(\alpha|\beta)$;

(c) $(\beta|\alpha) = \overline{(\alpha|\beta)}$, *onde a barra indica conjugação complexa*;

(d) $(\alpha|\alpha) > 0$ *se* $\alpha \neq 0$.

Deve-se observar que as condições (a), (b) e (c) implicam que

(e) $(\alpha|c\beta + \gamma) = \bar{c}(\alpha|\beta) + (\alpha|\gamma)$.

Outro fato merece ser mencionado. Quando F é o corpo R dos números reais, os complexos conjugados que aparecem em (c) e (e) são supérfluos; no entanto, no caso de F ser complexo eles são necessários para a consistência das condições. Sem estes complexos conjugados, teríamos a contradição:

$$(\alpha|\alpha) > 0 \quad \text{e} \quad (i\alpha|i\alpha) = -1(\alpha|\alpha) > 0.$$

Nos exemplos que seguem, como em todo o capítulo, F é o corpo dos números reais ou o corpo dos números complexos.

Exemplo 1. Sobre F^n existe um produto interno que denominamos o **produto interno canônico.** É definido sobre $\alpha = (x_1, \ldots, x_n)$ e $\beta = (y_1, \ldots, y_n)$ por

$$(\alpha|\beta) = \sum_j x_j \bar{y}_j. \tag{8-1}$$

Quando $F = R$, podemos também escrever

$$(\alpha|\beta) = \sum_j x_j y_j.$$

No caso real, o produto interno canônico é freqüentemente denominado produto escalar e é indicado por $\alpha \cdot \beta$.

Exemplo 2. Para $\alpha = (x_1, x_2)$ e $\beta = (y_1, y_2)$ em R^2, seja

$$(\alpha|\beta) = x_1y_1 - x_2y_1 - x_1y_2 + 4x_2y_2.$$

Como $(\alpha|\alpha) = (x_1 - x_2)^2 + 3x_2^2$, decorre que $(\alpha|\alpha) > 0$ se $\alpha \neq 0$. As condições (a), (b) e (c) da definição são facilmente verificadas.

Exemplo 3. Seja $V = F^{n \times n}$ o espaço de todas as $n \times n$ matrizes sobre F. Então V é isomorfo a F^{n2}, de uma maneira natural. Decorre, portanto, do Exemplo 1 que a equação

$$(A|B) = \sum_{j,k} A_{jk}\overline{B_{jk}}$$

define um produto interno sobre V. Além disso, introduzindo a matriz **transposta conjugada** B^*, onde $B^*_{kj} = \overline{B_{jk}}$, podemos exprimir este produto interno sobre $F^{n \times n}$ em termos da função traço

$$(A|B) = \operatorname{tr}(AB^*) = \operatorname{tr}(B^*A).$$

De fato,

$$\begin{aligned}\operatorname{tr}(AB^*) &= \sum_j (AB^*)_{jj}\\ &= \sum_j \sum_k A_{jk}B^*_{kj}\\ &= \sum_j \sum_k A_{jk}\overline{B_{jk}}.\end{aligned}$$

Exemplo 4. Seja $F^{n \times 1}$ o espaço das $n \times 1$ matrizes (colunas) sobre F e seja Q uma $n \times n$ matriz inversível sobre F. Para X, Y em $F^{n \times 1}$ definamos

$$(X|Y) = Y^*Q^*QX.$$

Estamos identificando a 1×1 matriz à direita com o seu único elemento. Quando Q é a matriz unidade, esse produto interno é essencialmente o mesmo que o do Exemplo 1; nós o denominamos **produto interno canônico** sobre $F^{n \times 1}$. O leitor deverá observar que a denominação "produto interno canônico" está sendo usada em dois contextos especiais. Para um espaço vetorial arbitrário sobre F, de dimensão finita, não existe nenhum produto interno evidente, que possa ser denominado canônico.

Exemplo 5. Seja V o espaço vetorial das funções contínuas definidas sobre o intervalo unitário, $0 \leq t \leq 1$ e tomando valores complexos. Seja

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\,dt.$$

O leitor provavelmente tem mais familiaridade com o espaço das funções contínuas definidas sobre o intervalo unitário e tomando valores reais, e para este espaço, a conjugação complexa sobre g pode ser omitida.

Exemplo 6. Este é na realidade toda uma classe de exemplos. Pode-se construir novos produtos internos a partir de um dado produto interno pelo seguinte método: Sejam V e W espaços vetoriais sobre F e suponhamos que $(\,|\,)$ seja um produto interno sobre W. Se T é uma transformação linear não-singular de V em W, então a equação

$$p_T(\alpha, \beta) = (T\alpha|T\beta)$$

define um produto interno p_T sobre V. O produto interno do Exemplo 4 é um caso particular desta situação. Os que seguem também são casos particulares.

(a) Seja V um espaço vetorial de dimensão finita e seja

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$$

uma base ordenada de V. Sejam $\varepsilon_1, \ldots, \varepsilon_n$ os vetores da base canônica de F^n e seja T a transformação linear de V em F^n tal que $T\alpha_j = \varepsilon_j$, $j = 1, \ldots, n$. Em outras palavras, seja T o isomorfismo "natural" de V em F^n determinado por $\mathscr{B}$. Se tomarmos o produto interno canônico sobre F^n, então

$$p_T(\sum_j x_j\alpha_j, \sum_k y_k\alpha_k) = \sum_{j=1}^{n} x_j\bar{y}_j.$$

Assim, para toda base $\mathscr{B}$ de V, existe um produto interno sobre V com a propriedade de que $(\alpha_j|\alpha_k) = \delta_{jk}$; na verdade, é fácil mostrar que existe exatamente um tal produto interno. Mostraremos posteriormente que todo produto interno sobre V é determinado por alguma base $\mathscr{B}$ da maneira acima.

(b) Consideremos novamente o Exemplo 5. Tomemos $V = W$, o espaço das funções contínuas sobre o intervalo unitário. Seja T o operador linear "multiplicação por t", isto é, $(Tf)(t) = tf(t)$, $0 \leq t \leq 1$. É fácil ver que T é linear. T também é não-singular; de fato, suponhamos que $Tf = 0$. Então, $tf(t) = 0$ para $0 \leq t \leq 1$; logo, $f(t) = 0$ para $t > 0$. Como f é contínua, temos também $f(0) = 0$ ou seja, $f = 0$. Usando o produto interno do Exemplo 5, construamos um novo produto interno sobre V colocando

$$p_T(f, g) = \int_0^1 (Tf)(t)\,\overline{(Tg)(t)}\,dt$$

$$= \int_0^1 f(t)\,\overline{g(t)}\,t^2 dt.$$

Voltemos agora a observações gerais sobre produtos internos. Suponhamos que V seja um espaço vetorial complexo com um produto interno. Então, para todos α, β em V

$$(\alpha|\beta) = Re(\alpha|\beta) + iIm(\alpha|\beta)$$

onde $Re(\alpha|\beta)$ e $Im(\alpha|\beta)$ são as partes real e imaginária do número complexo $(\alpha|\beta)$. Se z é um número complexo, então $Im(z) = Re(-iz)$. Decorre que

$$Im(\alpha|\beta) = Re[-i(\alpha|\beta)] = Re(\alpha|i\beta).$$

Assim, o produto interno é completamente determinado por sua "parte real" de acordo com

(8-2) $$(\alpha|\beta) = Re(\alpha|\beta) + iRe(\alpha|i\beta).$$

Às vezes é bastante útil saber que um produto interno sobre um espaço vetorial real ou complexo é determinado por uma outra função, a chamada **forma quadrática** determinada pelo produto interno. Para defini-la, indiquemos primeiro a raiz quadrada positiva de $(\alpha|\alpha)$ por $\|\alpha\|$; $\|\alpha\|$ é denominada a **norma** de α em relação ao produto interno. Observando os produtos internos canônicos em R^1, C^1, R^2 e R^3, o leitor poderá se convencer de que é conveniente considerar a norma de α como o "comprimento" ou "magnitude" de α. A **forma quadrática** determinada pelo produto interno é a

função que associa a cada vetor α o escalar $\|\alpha\|^2$. Decorre das propriedades do produto interno que

$$\|\alpha \pm \beta\|^2 = \|\alpha\|^2 \pm 2Re(\alpha|\beta) + \|\beta\|^2.$$

para todos os vetores α e β. Assim, no caso real,

(8-3) $$(\alpha|\beta) = \frac{1}{4}\|\alpha+\beta\|^2 - \frac{1}{4}\|\alpha-\beta\|^2.$$

No caso complexo usamos (8-2) para obter a expressão mais complicada:

(8-4) $$(\alpha|\beta) = \frac{1}{4}\|\alpha+\beta\|^2 - \frac{1}{4}\|\alpha-\beta\|^2 + \frac{i}{4}\|\alpha+i\beta\|^2 - \frac{i}{4}\|\alpha-i\beta\|^2.$$

As equações (8-3) e (8-4) são denominadas as **identidades de polarização.** Notemos que (8-4) pode também ser escrita como segue:

$$(\alpha|\beta) = \frac{1}{4}\sum_{n=1}^{4} i^n \|\alpha + i^n\beta\|^2.$$

As propriedades obtidas acima valem para qualquer produto interno sobre um espaço vetorial real ou complexo, não importando sua dimensão. Voltamos agora ao caso em que V é de dimensão finita. Como se pode imaginar, um produto interno sobre um espaço de dimensão finita sempre pode ser descrito em termos de uma base ordenada por meio de uma matriz.

Suponhamos então que V seja de dimensão finita, que

$$\mathscr{B} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$$

seja uma base ordenada de V e que nos seja dado um particular produto interno sobre V; mostraremos que ele é completamente determinado pelos valores

(8-5) $$G_{jk} = (\alpha_k|\alpha_j)$$

que assume sobre pares de vetores em $\mathscr{B}$. Se

$$\alpha = \sum_k x_k \alpha_k \quad \text{e} \quad \beta = \sum_j y_j \alpha_j$$

então

$$\begin{aligned}(\alpha|\beta) &= (\sum_k x_k \alpha_k | \beta) \\ &= \sum_k x_k (\alpha_k | \beta) \\ &= \sum_k x_k \sum_j \bar{y}_j (\alpha_k | \alpha_j) \\ &= \sum_{j,k} \bar{y}_j G_{jk} x_k \\ &= Y^* G X\end{aligned}$$

onde X, Y são as matrizes das coordenadas de α, β em relação à base ordenada $\mathscr{B}$ e G é a matriz com elementos $G_{jk} = (\alpha_k | \alpha_j)$. Denominamos G **a matriz do produto interno em relação à base ordenada** $\mathscr{B}$. Decorre de (8-5) que G é hermitiana, isto é, que $G = G^*$; contudo, G é uma matriz hermitiana especial. De fato, pois G deve satisfazer a condição adicional

(8-6) $$X^* G X > 0 \qquad X \neq 0.$$

Em particular, G deve ser inversível. Caso contrário, existiria uma $X \neq 0$ tal que $GX = 0$ e para uma tal X, (8-6) é impossível. Mais explicitamente, (8-6) afirma que para quaisquer escalares $x_1, \ldots, x_n$, não todos nulos

(8-7) $$\sum_{j,k} \bar{x}_j G_{jk} x_k > 0,$$

Daqui vemos imediatamente que todo elemento diagonal de G deve ser positivo; no entanto, esta condição sobre os elementos diagonais não é de forma alguma suficiente para assegurar a validez de (8-6). Condições suficientes para a validez de (8-6) serão dadas mais tarde.

O processo acima é reversível; isto é, se G é uma $n \times n$ matriz arbitrária sobre F que satisfaz (8-6), e a condição $G = G^*$, então G é a matriz, em relação à base ordenada $\mathscr{B}$, de algum produto interno sobre V. Tal produto interno é dado pela equação

$$(\alpha|\beta) = Y^*GX$$

onde X e Y são as matrizes das coordenadas de α e β em relação à base ordenada $\mathscr{B}$.

Exercícios

1. Seja V um espaço vetorial e $(|)$ um produto interno sobre V.

(a) Mostrar que $(0|\beta) = 0$ para todo β em V.

(b) Mostrar que se $(\alpha|\beta) = 0$ para todo β em V, então $\alpha = 0$.

2. Seja V um espaço vetorial sobre F. Mostrar que a soma de dois produtos internos sobre V é um produto interno sobre V. A diferença de dois produtos internos é um produto interno? Mostrar que um múltiplo positivo de um produto interno é um produto interno.

3. Descrever explicitamente todos os produtos internos sobre R^1 e sobre C^1.

4. Verificar que o produto interno canônico sobre F^n é um produto interno.

5. Seja $(\,|\,)$ o produto interno canônico sobre R^2.

(a) Sejam $\alpha = (1, 2)$, $\beta = (-1, 1)$. Se γ é um vetor tal que $(\alpha|\gamma) = -1$ e $(\beta|\gamma) = 3$, determinar γ.

(b) Mostrar que para todo α em R^2 temos $\alpha = (\alpha|\varepsilon_1)\,\varepsilon_1 + (\alpha|\varepsilon_2)\,\varepsilon_2$.

6. Seja $(\,|\,)$ o produto interno canônico sobre R^2 e seja T o operador linear $T(x_1, x_2) = (-x_2, x_1)$. Ora, T é "a rotação de 90°" e possui a propriedade de que $(\alpha|T\alpha) = 0$ para todo α em R^2. Determinar todos os produtos internos $[\,|\,]$ sobre R^2 tais que $[\alpha|T\alpha] = 0$ para todo α.

7. Seja $(\,|\,)$ o produto interno canônico sobre C^2. Demonstrar que não existe nenhum operador linear não-nulo sobre C^2 tal que $(\alpha|T\alpha) = 0$ para todo α em C^2. Generalizar.

8. Seja A uma 2×2 matriz com elementos reais. Para X, Y em $R^{2\times 1}$ seja

$$f_A(X, Y) = Y^t A X.$$

Mostrar que f_A é um produto interno sobre $A^{2\times 1}$ se, e somente se, $A = A^t$, $A_{11} > 0$, $A_{22} > 0$ e $\det A > 0$.

9. Seja V um espaço vetorial real ou complexo com produto interno. Mostrar que a forma quadrática determinada pelo produto interno satisfaz a **regra do paralelogramo**

$$\|\alpha + \beta\|^2 + \|\alpha - \beta\|^2 = 2\,\|\alpha\|^2 + 2\,\|\beta\|^2.$$

10. Seja $(\,|\,)$ o produto interno sobre R^2 definido no Exemplo 2 e seja $\mathcal{B}$ a base ordenada canônica de R^2. Determinar a matriz deste produto interno em relação a $\mathcal{B}$.

11. Mostrar que a fórmula

$$\left(\sum_j a_j x^j \,\middle|\, \sum_k b_k x^k\right) = \sum_{j,k} \frac{a_j b_k}{j + k + 1}$$

define um produto interno sobre o espaço $R[x]$ dos polinômios sobre o corpo R. Seja W o subespaço dos polinômios de grau menor ou igual a n. Restringir o produto interno acima a W e determinar a matriz deste produto interno sobre W em relação à base ordenada $\{1, x, x^2, \ldots, x^n\}$. (*Sugestão*: Para mostrar que a fórmula define um produto interno, observar que

$$(f|g) = \int_0^1 f(t)\, g(t)\, dt$$

e trabalhar com a integral.)

12. Seja V um espaço vetorial de dimensão finita e seja $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base de V. Seja $(\,|\,)$ um produto interno sobre V. Se $c_1, \ldots, c_n$ são n escalares arbitrários, mostrar que existe exatamente um vetor α em V tal que $(\alpha|\alpha_j) = c_j$, $j = 1, \ldots, n$.

13. Seja V um espaço vetorial complexo. Uma função J de V em V é denominada uma **conjugação** (também chamada função semilinear) se $J(\alpha + \beta) = J(\alpha) + J(\beta)$, $J(c\alpha) = \bar{c}J(\alpha)$ e $J(J(\alpha)) = \alpha$, para todos os escalares c e todos α, β em V. Se J é uma conjugação, mostrar que

(a) o conjunto W de todos α em V tais que $J\alpha = \alpha$ é um espaço vetorial sobre R em relação às operações definidas em V;

(b) para cada α em V existe um único par de vetores β, γ em W tais que $\alpha = \beta + i\gamma$.

14. Seja V um espaço vetorial complexo e W um subconjunto de V com as seguintes propriedades:

(a) W é um espaço vetorial real em relação às operações definidas em V.

(b) Para cada α em V existe um único par de vetores β, γ em W, tais que $\alpha = \beta + i\gamma$. Mostrar que a equação $J\alpha = \beta - i\gamma$ define uma conjugação sobre V tal que $J\alpha = \alpha$ se, e somente se, α pertence a W e mostrar também que J é a única conjugação sobre V com esta propriedade.

15. Determinar todas as conjugações sobre C^1 e C^2.

16. Seja W um subespaço real de dimensão finita de um espaço vetorial complexo V. Mostrar que W satisfaz a condição (b) do Exercício 14 se, e somente se, toda base de W é também uma base de V.

17. Seja V um espaço vetorial complexo, J uma conjugação sobre V, W o conjunto dos α em V tais que $J\alpha = \alpha$ e f um produto interno sobre W. Mostrar que:

(a) existe um único produto interno g sobre V tal que $g(\alpha, \beta) = f(\alpha, \beta)$ para todos α, β em W,

(b) $g(J\alpha, J\beta) = g(\beta, \alpha)$ para todos α, β em V.
O que a parte (a) diz acerca da relação entre os produtos internos canônicos sobre R^1 e C^1 ou sobre R^n e C^n?

8.2 Espaços com Produto Interno

Agora que temos alguma idéia sobre o que um produto interno é, voltaremos nossa atenção para o que pode ser dito a respeito da combinação de um espaço vetorial e algum produto interno particular sobre ele. Especificamente, estabeleceremos as propriedades básicas dos conceitos de "comprimento" e "ortogonalidade" que são impostas ao espaço pelo produto interno.

Definição. *Um* **espaço com produto interno** *é um espaço vetorial real ou complexo, munido de um produto interno especificado sobre aquele espaço.*

Um espaço real com produto interno e de dimensão finita é freqüentemente denominado um **espaço euclidiano.** Um espaço complexo com produto interno é freqüentemente dito um **espaço unitário.**

Teorema 1. *Se* V *é um espaço com produto interno, então, para quaisquer vetores* α, β *em* V *e todo escalar* c

(i) $\|c\alpha\| = |c| \, \|\alpha\|$;
(ii) $\|\alpha\| > 0$ *para* $\alpha \neq 0$;
(iii) $|(\alpha|\beta)| \leq \|\alpha\| \, \|\beta\|$;
(iv) $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$.

Demonstração. As afirmações (i) e (ii) decorrem quase imediatamente das diversas definições envolvidas. A desigualdade em (iii) é evidentemente válida quando $\alpha = 0$. Se $\alpha \neq 0$, coloquemos

$$\gamma = \beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha.$$

Então $(\gamma|\alpha) = 0$ e

$$0 \leq \|\gamma\|^2 = \left(\beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha \middle| \beta - \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha\right)$$

$$= (\beta|\beta) - \frac{(\beta|\alpha)(\alpha|\beta)}{\|\alpha\|^2}$$

$$= \|\beta\|^2 - \frac{|(\alpha|\beta)|^2}{\|\alpha\|^2}.$$

Logo, $|(\alpha|\beta)|^2 \leq \|\alpha\|^2 \|\beta\|^2$. Usando agora (iii) concluímos que

$$\begin{aligned} \|\alpha + \beta\|^2 &= \|\alpha\|^2 + (\alpha|\beta) + (\beta|\alpha) + \|\beta\|^2 \\ &= \|\alpha\|^2 + 2Re(\alpha|\beta) + \|\beta\|^2 \\ &\leq \|\alpha\|^2 + 2\|\alpha\| \|\beta\| + \|\beta\|^2 \\ &= (\|\alpha\| + \|\beta\|)^2. \end{aligned}$$

Assim, $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$.

A desigualdade em (iii) é denominada a **desigualdade de Cauchy-Schwarz.** Ela possui uma ampla gama de aplicações. A demonstração mostra (por exemplo) que se α for não-nulo, então $|(\alpha|\beta)| < \|\alpha\| \|\beta\|$ a não ser que

$$\beta = \frac{(\beta|\alpha)}{\|\alpha\|^2}\alpha.$$

Portanto, a igualdade ocorre em (iii) se, e somente se, α e β são linearmente dependentes.

Exemplo 7. Aplicando a desigualdade de Cauchy-Schwarz aos produtos internos dados nos Exemplos 1, 2, 3 e 5 obtemos o seguinte:

(a) $$|\Sigma x_k \bar{y}_k| \leq (\Sigma |x_k|^2)^{1/2} (\Sigma |y_k|^2)^{1/2}$$

(b) $$|x_1 y_1 - x_2 y_1 - x_1 y_2 + 4x_2 y_2| \leq ((x_1 - x_2)^2 + 3x_2^2)^{1/2}((y_1 - y_2)^2 + 3y_2^2)^{1/2}$$

(c) $$|\text{tr}(AB^*)| \leq (\text{tr}(AA^*))^{1/2} (\text{tr}(BB^*))^{1/2}$$

(d) $$\left| \int_0^1 f(x)\overline{g(x)}dx \right| \leq \left(\int_0^1 |f(x)|^2 dx \right)^{1/2} \left(\int_0^1 |g(x)|^2 dx \right)^{1/2}.$$

Definições. *Sejam α e β vetores num espaço V com produto interno. Dizemos que α é ortogonal a β se $(\alpha|\beta) = 0$; como isso implica que β é ortogonal a α, dizemos simplesmente que α e β são ortogonais. Se S é um conjunto de vetores em V, dizemos que S é um*

conjunto ortogonal *se dois quaisquer vetores distintos em* S *são ortogonais. Um* **conjunto ortonormal** *é um conjunto ortogonal* S, *com a propriedade adicional de que* $\|\alpha\| = 1$ *para todo* α *em* S.

O vetor nulo é ortogonal a todo vetor em V **e** é o único vetor com esta propriedade. É conveniente pensar em um conjunto ortonormal como um conjunto de vetores mutuamente perpendiculares, cada um tendo comprimento 1.

Exemplo 8. A base canônica de R^n ou de C^n é um conjunto ortonormal em relação ao produto interno canônico.

Exemplo 9. O vetor (x, y) em R^2 é ortogonal a $(-y, x)$ em relação ao produto interno canônico, pois

$$((x, y)|(-y, x)) = -xy + yx = 0.$$

No entanto, se R^2 estiver munido do produto interno do Exemplo 2, (x, y) e $(-y, x)$ serão ortogonais se, e somente se,

$$y = \frac{1}{2}(-3 \pm \sqrt{13})\,x.$$

Exemplo 10. Seja $V = C^{n \times n}$ o espaço das $n \times n$ matrizes complexas e E^{pq} a matriz cujo único elemento não-nulo é um 1 na linha p e coluna q. Então, o conjunto de todas estas matrizes E^{pq} é ortonormal em relação ao produto interno dado no Exemplo 3. De fato, pois

$$(E^{pq}|E^{rs}) = \operatorname{tr}(E^{pq}E^{sr}) = \delta_{qs}\operatorname{tr}(E^{pr}) = \delta_{qs}\,\delta_{pr}.$$

Exemplo 11. Se V é o espaço das funções contínuas definidas sobre o intervalo $0 \leq x \leq 1$ e tomando valores complexos (ou valores reais), com o produto interno

$$(f|g) = \int_0^1 f(x)\overline{g(x)}\,dx$$

Suponhamos $f_n(x) = \sqrt{2}\cos 2\pi nx$ e $g_n(x) = \sqrt{2}\operatorname{sen} 2\pi nx$. Então $\{1, f_1, g_1, f_2, g_2, \ldots\}$ é um conjunto ortonormal infinito. No caso complexo, podemos também formar as combinações lineares

$$\frac{1}{\sqrt{2}}(f_n + ig_n), \qquad n = 1, 2, \ldots$$

Dessa maneira obteremos um novo conjunto ortonormal S constituído de todas as funções da forma

$$h_n(x) = e^{2\pi inx}, \qquad n = \pm 1, \ \pm 2, \ \ldots$$

Acrescentando-se ao conjunto S a função constante 1 obtém-se o conjunto S' que também é ortonormal. Estamos supondo aqui que o leitor esteja familiarizado com o cálculo das integrais em questão.

Os conjuntos ortonormais dados nos exemplos acima são todos linearmente independentes. Mostraremos agora que necessariamente isso acontece.

Teorema 2. *Um conjunto ortogonal de vetores não-nulos é linearmente independente.*

Demonstração. Seja S um conjunto ortogonal finito ou infinito de vetores não-nulos num espaço com produto interno. Suponhamos que $\alpha_1, \alpha_2, \ldots, \alpha_m$ sejam vetores distintos em S e que

$$\beta = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_m\alpha_m.$$

Então

$$\begin{aligned}(\beta|\alpha_k) &= \sum_j (c_j\alpha_j|\alpha_k) \\ &= \sum_j c_j(\alpha_j|\alpha_k) \\ &= c_k(\alpha_k|\alpha_k).\end{aligned}$$

Como $(\alpha_k|\alpha_k) \neq 0$, decorre que

$$c_k = \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2}, \quad 1 \leq k \leq m.$$

Assim, quando $\beta = 0$, cada $c_k = 0$; logo S é um conjunto independente.

Corolário. *Se um vetor β é uma combinação linear de uma seqüência ortogonal de vetores não-nulos $\alpha_1, \alpha_2, \ldots, \alpha_m$ então β é exatamente a combinação linear*

(8-8) $$\beta = \sum_{k=1}^{m} \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2} \alpha_k.$$

Este corolário decorre da demonstração do teorema. Existe um outro corolário que, apesar de evidente, deve ser mencionado. Se $\{\alpha_1, \ldots, \alpha_m\}$ é um conjunto ortogonal de vetores não-nulos em um espaço V de dimensão finita com produto interno, então $m < \dim V$. Isso diz que o número de direções mutuamente ortogonais em V não pode exceder a dimensão de V, algebricamente definida. O número máximo de direções mutuamente ortogonais em V é o que, intuitivamente, se consideraria como sendo a dimensão geométrica de V, e acabamos de ver que esta não é maior que a dimensão algébrica. O fato de que essas duas dimensões sejam iguais, é um corolário particular do próximo teorema.

Teorema 3. *Seja* V *um espaço com produto interno e sejam* $\beta_1, \ldots, \beta_n$ *vetores independentes arbitrários em* V. *Então é possível construir vetores ortogonais* $\alpha_1, \ldots, \alpha_n$ *em* V *tais que, para cada* $k = 1, 2, \ldots, n$, *o conjunto*

$$\{\alpha_1, \ldots, \alpha_k\}$$

é uma base do subespaço gerado por $\beta_1, \ldots, \beta_k$.

Demonstração. Os vetores $\alpha_1, \ldots, \alpha_n$ serão obtidos por meio de uma construção conhecida como o **processo de ortogonalização de Gram-Schmidt.** Inicialmente, seja $\alpha_1 = \beta_1$. Os outros vetores serão obtidos, indutivamente, da seguinte maneira: Suponhamos que $\alpha_1, \ldots, \alpha_m (1 \leq m < n)$ tenham sido escolhidos de tal modo que, para cada k

$$\{\alpha_1, \ldots, \alpha_k\}, \qquad 1 \leq k \leq m,$$

seja uma base ortogonal para o subespaço de V, gerado por $\beta_1, \ldots, \beta_k$. Para construir o vetor seguinte, α_{m+1}, seja

(8-9) $$\alpha_{m+1} = \beta_{m+1} - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{\|\alpha_k\|^2} \alpha_k.$$

Então $\alpha_{m+1} \neq 0$ pois, caso contrário, β_{m+1} seria uma combinação linear de $\alpha_1, \ldots, \alpha_m$ e, portanto, uma combinação linear de $\beta_1, \ldots, \beta_m$.

Além do mais, se $1 \leq j \leq m$,

$$\begin{aligned}(\alpha_{m+1}|\alpha_j) &= (\beta_{m+1}|\alpha_j) - \sum_{k=1}^{m} \frac{(\beta_{m+1}|\alpha_k)}{\|\alpha_k\|^2}(\alpha_k|\alpha_j)\\ &= (\beta_{m+1}|\alpha_j) - (\beta_{m+1}|\alpha_j)\\ &= 0.\end{aligned}$$

Portanto, $\{\alpha_1, \ldots, \alpha_{m+1}\}$ é um conjunto ortogonal formado de $m+1$ vetores não nulos do subespaço gerado por $\beta_1, \ldots, \beta_{m+1}$. Pelo Teorema 2, ele constitui uma base desse subespaço. Assim, os vetores $\alpha_1, \ldots, \alpha_n$ podem ser construídos, um após o outro, de acordo com (8-9). Em particular, se $n = 4$, temos

$$\begin{aligned}\alpha_1 &= \beta_1\\ \alpha_2 &= \beta_2 - \frac{(\beta_2|\alpha_1)}{\|\alpha_1\|^2}\alpha_1\\ \alpha_3 &= \beta_3 - \frac{(\beta_3|\alpha_1)}{\|\alpha_1\|^2}\alpha_1 - \frac{(\beta_3|\alpha_2)}{\|\alpha_2\|^2}\alpha_2\\ \alpha_4 &= \beta_4 - \frac{(\beta_4|\alpha_1)}{\|\alpha_1\|^2}\alpha_1 - \frac{(\beta_4|\alpha_2)}{\|\alpha_2\|^2}\alpha_2 - \frac{(\beta_4|\alpha_3)}{\|\alpha_3\|^2}\alpha_3.\end{aligned} \tag{8-10}$$

Corolário. *Todo espaço de dimensão finita com produto interno possui uma base ortonormal.*

Demonstração. Seja V um espaço de dimensão finita com produto interno e seja $\{\beta_1, \ldots, \beta_n\}$ uma base de V. Apliquemos o processo de Gram-Schmidt para construir uma base ortogonal $\{\alpha_1, \ldots, \alpha_n\}$. Para obter em seguida uma base ortonormal, substituamos simplesmente cada vetor α_k por $\alpha_k/\|\alpha_k\|$.

Uma das principais vantagens que as bases ortonormais possuem, em comparação com bases arbitrárias, é que os cálculos envolvendo coordenadas tornam-se mais simples. Para mostrar, em termos gerais, por que isso é verdade, suponhamos que V seja um espaço de dimensão finita com produto interno. Podemos então, como na última seção, usar a Equação (8-5) para associar a cada base ordenada $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ de V, uma matriz G. Usando essa matriz

$$G_{jk} = (\alpha_k|\alpha_j),$$

podemos calcular produtos internos em termos de coordenadas. Se $\mathscr{B}$ for uma base ortonormal, G será a matriz identidade e para quaisquer escalares x_j e y_k

$$(\sum_j x_j\alpha_j|\sum_k y_k\alpha_k) = \sum_j x_j\bar{y}_j.$$

Assim, em termos de uma base ortonormal, o produto interno sobre V tem a mesma aparência que o produto interno canônico sobre F^n.

Apesar do seu limitado uso prático em cálculos, é interessante observar que o processo de Gram-Schmidt pode também ser usado para testar dependência linear. De fato, suponhamos que $\beta_1, \ldots, \beta_n$ sejam vetores linearmente dependentes em um espaço V com produto interno. Para excluirmos um caso trivial, suponhamos $\beta_1 \neq 0$. Seja m o maior inteiro para o qual os vetores $\beta_1, \ldots, \beta_m$ são independentes. Então $1 \leq m < n$. Sejam $\alpha_1, \ldots, \alpha_n$ os vetores obtidos, aplicando-se aos vetores $\beta_1, \ldots, \beta_m$ o processo de ortogonalização. Então o vetor α_{m+1}, dado por (8-9), é necessariamente nulo. De fato, α_{m+1} está no subespaço gerado por $\alpha_1, \ldots, \alpha_m$ e é ortogonal a cada um desses vetores; logo α_{m+1} é nulo por (8-8). Reciprocamente, se $\alpha_1, \ldots, \alpha_m$ são não nulos e $\alpha_{m+1} = 0$, os vetores $\beta_1, \ldots, \beta_{m+1}$ são linearmente dependentes.

Exemplo 12. Consideremos os vetores

$$\begin{aligned}\beta_1 &= (3, 0, 4)\\ \beta_2 &= (-1, 0, 7)\\ \beta_3 &= (2, 9, 11)\end{aligned}$$

em R^3 munido do produto interno canônico. Aplicando aos vetores β_1, β_2, β_3 o processo de Gram-Schmidt, obteremos os seguintes vetores

$$\alpha_1 = (3, 0, 4)$$

$$\begin{aligned}\alpha_2 &= (-1, 0, 7) - \frac{((-1, 0, 7)|(3, 0, 4))}{25}(3, 0, 4)\\ &= (-1, 0, 7) - (3, 0, 4)\\ &= (-4, 0, 3)\end{aligned}$$

$$\begin{aligned}\alpha_3 &= (2, 9, 11) - \frac{((2, 9, 11)|(3, 0, 4))}{25}(3, 0, 4)\\ &\quad - \frac{((2, 9, 11)|(-4, 0, 3))}{25}(-4, 0, 3)\\ &= (2, 9, 11) - 2(3, 0, 4) - (-4, 0, 3)\\ &= (0, 9, 0).\end{aligned}$$

Esses vetores são obviamente não nulos e mutuamente ortogonais. Logo, $\{\alpha_1, \alpha_2, \alpha_3\}$ é uma base ortogonal de R^3. Para exprimir um vetor arbitrário (x_1, x_2, x_3) de R^3, como uma combinação linear de $\alpha_1, \alpha_2, \alpha_3$ *não* é necessário resolver nenhuma equação linear. Basta usar (8-8). Assim,

$$(x_1, x_2, x_3) = \frac{3x_1 + 4x_3}{25}\alpha_1 + \frac{-4x_1 + 3x_3}{25}\alpha_2 + \frac{x_2}{9}\alpha_3$$

como é facilmente verificável. Em particular,

$$(1, 2, 3) = \tfrac{3}{5}(3, 0, 4) + \tfrac{1}{5}(-4, 0, 3) + \tfrac{2}{9}(0, 9, 0).$$

Sob outro aspecto, o que mostramos foi o seguinte: A base $\{f_1, f_2, f_3\}$ de $(R^3)^*$, dual da base $\{\alpha_1, \alpha_2, \alpha_3\}$ é definida explicitamente pelas equações

$$f_1(x_1, x_2, x_3) = \frac{3x_1 + 4x_3}{25}$$

$$f_2(x_1, x_2, x_3) = \frac{-4x_1 + 3x_3}{25}$$

$$f_3(x_1, x_2, x_3) = \frac{x^2}{9}$$

e essas equações podem ser escritas, mais geralmente, na forma

$$f_j(x_1, x_2, x_3) = \frac{((x_1, x_2, x_3)|\alpha_j)}{\|\alpha_j\|^2}.$$

Finalmente, observemos que de $\alpha_1, \alpha_2, \alpha_3$ obtemos a base ortonormal

$$\{\tfrac{1}{5}(3, 0, 4),\ \tfrac{1}{5}(-4, 0, 3),\ (0, 1, 0)\}.$$

Exemplo 13. Seja $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ onde a, b, c e d são números complexos. Coloquemos $\beta_1 = (a, b)$, $\beta_2 = (c, d)$ e suponhamos que $\beta_1 \neq 0$. Se aplicarmos o processo de ortogonalização a β_1, β_2 usando o produto interno canônico de C^2, obteremos os seguintes vetores

$$\alpha_1 = (a, b)$$

$$\begin{aligned}\alpha_2 &= (c, d) - \frac{((c, d)|(a, b))}{|a|^2 + |b|^2}(a, b) \\ &= (c, d) - \frac{(c\bar{a} + d\bar{b})}{|a|^2 + |b|^2}(a, b) \\ &= \left(\frac{cb\bar{b} - d\bar{b}a}{|a|^2 + |b|^2}, \frac{d\bar{a}a - c\bar{a}b}{|a|^2 + |b|^2}\right) \\ &= \frac{\det A}{|a|^2 + |b|^2}(-\bar{b}, \bar{a}).\end{aligned}$$

Ora, a teoria geral nos diz que $\alpha_2 \neq 0$ se, e somente se, β_1 e β_2 forem linearmente independentes. Por outro lado, a fórmula para α_2 mostra que isso acontece se, e somente se, $\det A \neq 0$.

Em essência, o processo de Gram-Schmidt consiste de aplicações repetidas de uma operação geométrica fundamental, denominada projeção ortogonal, e o processo é melhor compreendido sob este ponto de vista. O método da projeção ortogonal também surge, de um modo natural, na solução de problemas importantes de aproximação.

Suponhamos que W seja um subespaço de um espaço com produto interno V e β um vetor arbitrário de V. O problema consiste em determinar uma melhor aproximação possível de β por vetores de W. Isso significa que devemos determinar um vetor α para o qual $\|\beta - \alpha\|$ seja tão pequeno quanto possível e sujeito à restrição de que α seja um elemento de W. Precisemos nossa linguagem.

Uma **melhor aproximação** de β por vetores de W é um vetor α em W tal que

$$\|\beta - \alpha\| \leq \|\beta - \gamma\|$$

para todo vetor γ em W.

Examinando esse problema em R^2 ou em R^3 vemos, intuitivamente, que uma melhor aproximação de β por vetores de W deverá ser um vetor α em W tal que $\beta - \alpha$ seja perpendicular (ortogonal) a W e que deverá existir exatamente um α nessas condições. Essas idéias intuitivas são corretas para subespaços de dimensão finita e para alguns, mas não todos, subespaços de dimensão infinita. Como a situação precisa é demasiadamente complicada para ser tratada aqui, demonstraremos apenas o seguinte resultado:

Teorema 4. *Seja* W *um subespaço de um espaço* V *com produto interno e seja* β *um vetor de* V.

(i) *O vetor* α *em* W *é uma melhor aproximação de* β *por vetores de* W *se, e somente se,* $\beta - \alpha$ *é ortogonal a todo vetor de* W.

(ii) *Se uma melhor aproximação de* β *por vetores de* W *existir, ela é única.*

(iii) *Se* W *é de dimensão finita e* $\{\alpha_1, \ldots, \alpha_n\}$ *é uma base ortonormal arbitrária de* W, *o vetor*

$$\alpha = \sum_k \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2}\alpha_k$$

é a (única) melhor aproximação de β *por vetores de* W.

Demonstração. Inicialmente observemos que se γ é um vetor arbitrário de V, então $\beta - \gamma = (\beta - \alpha) + (\alpha - \gamma)$, e

$$\|\beta - \gamma\|^2 = \|\beta - \alpha\|^2 + 2\,\mathrm{Re}(\beta - \alpha|\alpha - \gamma) + \|\alpha - \gamma\|^2.$$

Suponhamos agora que $\beta - \alpha$ seja ortogonal a todo vetor em W, que γ esteja em W e que $\gamma \neq \alpha$. Como $\alpha - \gamma$ está em W, segue-se que

$$\begin{aligned}\|\beta - \gamma\|^2 &= \|\beta - \alpha\|^2 + \|\alpha - \gamma\|^2 \\ &> \|\beta - \alpha\|^2.\end{aligned}$$

Reciprocamente, suponhamos que $\|\beta - \gamma\| \geq \|\beta - \alpha\|$ para todo γ em W. Então, segue-se, da primeira equação acima, que

$$2\,\mathrm{Re}(\beta - \alpha|\alpha - \gamma) + \|\alpha - \gamma\|^2 \geq 0$$

para todo γ em W. Como todo vetor em W pode ser expresso na forma $\alpha - \gamma$ com γ em W, vemos que

$$2\,\mathrm{Re}(\beta - \alpha|\tau) + \|\tau\|^2 \geq 0$$

para todo r em W. Em particular, se γ está em W e $\gamma \neq \alpha$, podemos colocar

$$\tau = -\frac{(\beta - \alpha|\alpha - \gamma)}{\|\alpha - \gamma\|^2}(\alpha - \gamma).$$

Então a desigualdade se reduz a

$$-2\frac{|(\beta - \alpha|\alpha - \gamma)|^2}{\|\alpha - \gamma\|^2} + \frac{|(\beta - \alpha|\alpha - \gamma)|^2}{\|\alpha - \gamma\|^2} \geq 0.$$

Essa expressão é válida se, e somente se, $(\beta - \alpha|\alpha - \gamma) = 0$. Portanto, $\beta - \alpha$ é ortogonal a todo vetor de W. Isso completa a demonstração da equivalência das duas condições sobre α, dadas em (i). A condição de ortogonalidade é evidentemente satisfeita por, no máximo, um vetor de W, o que demonstra (ii).

Suponhamos agora que W seja um subespaço de dimensão finita de V. Então, como corolário do Teorema 3, sabemos que W possui uma base ortogonal. Seja $\{\alpha_1, \ldots, \alpha_n\}$ uma base ortogonal arbitrária de W e definamos α por (8-11). Então, em vista do cálculo feito na demonstração do Teorema 3, $\beta - \alpha$ é ortogonal a cada um dos vetores α_k($\beta - \alpha$ é o vetor obtido na última passagem, quando o processo de ortogonalização é aplicado a $\alpha_1, \ldots, \alpha_n, \beta$). Assim, $\beta - \alpha$ é ortogonal a toda combinação linear de $\alpha_1, \ldots, \alpha_n$, isto é, a todo vetor de W. Se γ está em W e $\gamma \neq \alpha$, segue-se que $\|\beta - \gamma\| > \|\beta - \alpha\|$. Portanto, α é a melhor aproximação de β contida em W.

Definição. *Seja* V *um espaço com produto interno e seja* S *um conjunto arbitrário de vetores em* V. *O* **suplementar ortogonal** *de* S *é o conjunto* $S^\perp$ *de todos os vetores em* V *que são ortogonais a todo vetor em* S.

O suplementar ortogonal de V é o subespaço nulo e, reciprocamente $\{0\}^\perp = V$. Se S é um subconjunto arbitrário de V, seu suplementar ortogonal $S^\perp$ (S perpendicular) é sempre um subespaço de V. De fato, $S^\perp$ é não vazio pois ele contém 0; e sempre que α e β estiverem em $S^\perp$ e c for um escalar arbitrário,

$$\begin{aligned}(c\alpha + \beta|\gamma) &= c(\alpha|\gamma) + (\beta|\gamma)\\ &= c0 + 0\\ &= 0\end{aligned}$$

para todo γ em S, logo $c\alpha + \beta$ também estará em $S^{\perp}$. No Teorema 4, a propriedade característica do vetor α é a de que ele é o único vetor em W tal que $\beta - \alpha$ pertença a $W^{\perp}$.

Definição. *O vetor* α *do Teorema 4, sempre que existir, é denominado* **projeção ortogonal de β sobre W.** *Se todo vetor de* V *tiver uma projeção ortogonal sobre* W, *a aplicação que associa a cada vetor de* V *sua projeção ortogonal sobre* W *é denominada* **projeção ortogonal de V sobre W.**

Pelo Teorema 4, sempre existe a projeção ortogonal de um espaço com produto interno, sobre um subespaço de dimensão finita. Mas o Teorema 4 também implica o seguinte resultado:

Corolário. *Sejam* V *um espaço com produto interno,* W *um subespaço de dimensão finita e* E *a projeção ortogonal de* V *sobre* W. *Então a aplicação*

$$\beta \to \beta - E\beta$$

é a projeção ortogonal de V *sobre* $W^{\perp}$.

Demonstração. Seja β um vetor arbitrário em V. Então $\beta - E\beta$ está em $W^{\perp}$, e para qualquer γ em $W^{\perp}$, $\beta - \gamma = E\beta + (\beta - E\beta - \gamma)$. Como $E\beta$ está em W e $\beta - E\beta - \gamma$ está em $W^{\perp}$, segue-se que

$$\begin{aligned} \|\beta - \gamma\|^2 &= \|E\beta\|^2 + \|\beta - E\beta - \gamma\|^2 \\ &\geq \|\beta - (\beta - E\beta)\|^2 \end{aligned}$$

valendo a desigualdade estrita quando $\gamma \neq \beta - E\beta$. Portanto, $\beta - E\beta$ é a melhor aproximação de β por vetores em $W^{\perp}$.

Exemplo 14. Seja dado R^3, munido do produto interno canônico. A projeção ortogonal de $(-10, 2, 8)$ sobre o subespaço W gerado por $(3, 12, -1)$ é o vetor

$$\alpha = \frac{((-10, 2, 8)|(3, 12, -1))}{9 + 144 + 1}(3, 12, -1)$$

$$= \frac{-14}{154}(3, 12, -1).$$

A projeção ortogonal de R^3 sobre W é a transformação linear E, definida por

$$(x_1, x_2, x_3) \to \left(\frac{3x_1 + 12x_2 - x_3}{154}\right)(3, 12, -1).$$

O posto de E é, evidentemente, 1; portanto, sua nulidade é 2. Por outro lado,

$$E(x_1, x_2, x_3) = (0, 0, 0)$$

se, e somente se, $3x_1 + 12x_2 - x_3 = 0$. Isso acontece se, e somente se, (x_1, x_2, x_3) estiver em $W^\perp$. Portanto, $W^\perp$ é o núcleo de E e $\dim(W^\perp) = 2$. Calculando

$$(x_1, x_2, x_3) - \left(\frac{3x_1 + 12x_2 - x_3}{154}\right)(3, 12, -1)$$

vemos que a projeção ortogonal de R^3 sobre $W^\perp$ é a transformação linear $I - E$ que leva o vetor (x_1, x_2, x_3) no vetor

$$\frac{1}{154}(145x_1 - 36x_2 + 3x_3, -36x_1 + 10x_2 + 12x_3, 3x_1 + 12x_2 + 153x_3).$$

As observações feitas no Exemplo 14 generalizam-se da seguinte maneira:

Teorema 5. *Seja* W *um subespaço de dimensão finita de um espaço* V *com produto interno e seja* E *a projeção ortogonal de* V *sobre* W. *Então* E *é uma transformação linear idempotente de* V *sobre* W, $\mathrm{W}^\perp$ *é o núcleo de* E *e*

$$\mathrm{V} = \mathrm{W} \oplus \mathrm{W}^\perp.$$

Demonstração. Seja β um vetor arbitrário de V. Então $E\beta$ é a melhor aproximação de β contida em W. Em particular, $E\beta = \beta$ quando β estiver em W. Portanto, $E(E\beta) = E\beta$ para todo β em V; isto é, E é idempotente: $E^2 = E$. Para provar que E é uma transformação linear, sejam α e β vetores quaisquer em V e c um escalar arbitrário. Pelo Teorema 4, $\alpha - E\alpha$ e $\beta - E\beta$ são ortogonais a todos os vetores de W. Portanto, o vetor

$$c(\alpha - E\alpha) + (\beta - E\beta) = (c\alpha + \beta) - (cE\alpha + E\beta)$$

também pertence a $W^{\perp}$. Como $cE\alpha + E\beta$ é um vetor em W, segue-se, do Teorema 4, que

$$E(c\alpha + \beta) = cE\alpha + E\beta.$$

É claro que podemos também demonstrar a linearidade de E usando (8-11). Novamente seja β um vetor arbitrário em V. Então $E\beta$ é o único vetor em W tal que $\beta - E\beta$ esteja em $W^{\perp}$. Portanto, $E\beta = 0$ quando β estiver em $W^{\perp}$. Reciprocamente, β estará em $W^{\perp}$ quando $E\beta = 0$. Logo $W^{\perp}$ é o núcleo de E. A equação

$$\beta = E\beta + \beta - E\beta$$

mostra que $V = W + W^{\perp}$; além disso, $W \cap W^{\perp} = \{0\}$. De fato, se α é um vetor em $W \cap W^{\perp}$, então $(\alpha|\alpha) = 0$. Portanto, $\alpha = 0$ e V é a soma direta de W e $W^{\perp}$.

Corolário. *Nas condições do teorema,* I − E *é a projeção ortogonal de* V *sobre* $W^{\perp}$. *Ela é uma transformação linear idempotente de* V *sobre* $W^{\perp}$ *com núcleo* W.

Demonstração. Já vimos que a aplicação $\beta \to \beta - E\beta$ é a projeção ortogonal de V sobre $W^{\perp}$. Como E é uma transformação linear, essa projeção sobre $W^{\perp}$ é a transformação linear $I - E$. Das suas propriedades geométricas, vê-se que $I - E$ é uma transformação idempotente de V sobre W. Isso decorre também do cálculo

$$\begin{aligned}(I - E)(I - E) &= I - E - E + E^2 \\ &= I - E.\end{aligned}$$

Além disso, $(I - E)\beta = 0$ se, e somente se, $\beta = E\beta$, e isso acontece se, e somente se, β estiver em W. Portanto, W é o núcleo de $I - E$.

O processo de Gram-Schmidt pode, agora, ser descrito geometricamente, da seguinte maneira: Dados um espaço V com produto interno e vetores $\beta_1, \ldots, \beta_n$ em V, seja $P_k (k > 1)$ a projeção ortogonal de V sobre o suplementar ortogonal do subespaço gerado por $\beta_1, \ldots, \beta_{k-1}$, e coloquemos $P_1 = I$. Então os vetores que se obtém, aplicando o processo de ortogonalização a $\beta_1, \ldots, \beta_n$, são definidos pelas equações

(8-12) $$\alpha_k = P_k\beta_k, \qquad 1 \leq k \leq n.$$

O Teorema 5 implica um outro resultado conhecido como a **desigualdade de Bessel.**

Corolário. *Seja* $\{\alpha_1, \ldots, \alpha_n\}$ *um conjunto ortogonal de vetores não-nulos* ***em*** *um espaço* V *com produto interno. Se* β *é um vetor arbitrário em* V, *então*

$$\sum_k \frac{|(\beta|\alpha_k)|^2}{\|\alpha_k\|^2} \leq \|\beta\|^2$$

e a igualdade vale se, e somente se,

$$\beta = \sum_k \frac{(\beta|\alpha_k)}{\|\alpha_k\|^2} \alpha_k.$$

Demonstração. Seja $\gamma = \sum_k [(\beta|\alpha_k)/\|\alpha_k\|^2]\alpha_k$. Então $\beta = \gamma + \delta$, onde $(\gamma|\delta) = 0$. Logo,

$$\|\beta\|^2 = \|\gamma\|^2 + \|\delta\|^2.$$

Agora, basta demonstrar que

$$\|\gamma\|^2 = \sum_k \frac{|(\beta|\alpha_k)|^2}{\|\alpha_k\|^2}.$$

Isto é um cálculo imediato, usando o fato de que $(\alpha_j|\alpha_k) = 0$ para $j \neq k$.

No caso especial em que $\{\alpha_1, \ldots, \alpha_n\}$ é um conjunto ortonormal, a desigualdade de Bessel afirma que

$$\sum_k |(\beta|\alpha_k)|^2 \leq \|\beta\|^2.$$

O corolário também nos diz que, neste caso, β está no subespaço gerado por $\alpha_1, \ldots, \alpha_n$ se, e somente se,

$$\beta = \sum_k (\beta|\alpha_k)\,\alpha_k.$$

ou seja, se, e somente se, a desigualdade de Bessel é, na verdade, uma igualdade. É claro que, no caso de V ser de dimensão finita e $\{\alpha_1, \ldots, \alpha_n\}$ ser uma base ortonormal de V, a fórmula acima vale

para todo vetor β em V. Em outras palavras, se $\{\alpha_1, \ldots, \alpha_n\}$ é uma base ortonormal de V, a k-ésima coordenada de β em relação à base ordenada $\{\alpha_1, \ldots, \alpha_n\}$ é $(\beta|\alpha_k)$.

Exemplo 15. Aplicaremos o último corolário aos conjuntos ortonormais descritos no Exemplo 11. Concluímos que

(a) $$\sum_{k=-n}^{n} \left| \int_0^1 f(t) e^{-2\pi i k t}\, dt \right|^2 \leq \int_0^1 |f(t)|^2\, dt$$

(b) $$\int_0^1 \left| \sum_{k=-n}^{n} c_k e^{2\pi i k t} \right|^2 dt = \sum_{k=-n}^{n} |c_k|^2$$

(c) $$\int_0^1 (\sqrt{2}\cos 2\pi t + \sqrt{2}\,\text{sen}\, 4\pi t)^2\, dt = 1 + 1 = 2.$$

Exercícios

1. Consideremos R^4 com o produto interno canônico. Seja W o subespaço de R^4 formado pelos vetores que são ortogonais a $\alpha = (1, 0, -1, 1)$ e a $\beta = (2, 3, -1, 2)$. Determinar uma base de W.

2. Aplicar o processo de Gram-Schmidt aos vetores $\beta_1 = (1, 0, 1)$, $\beta_2 = (1, 0, -1)$, $\beta_3 = (0, 3, 4)$, para obter uma base ortonormal de R^3 com o produto interno canônico.

3. Consideremos C^3 com o produto interno canônico. Determinar uma base ortonormal do subespaço gerado por $\beta_1 = (1, 0, i)$ e $\beta_2 = (2, 1, 1 + i)$.

4. Seja V um espaço com produto interno. A distância entre dois vetores α e β em V é definida por

$$d(\alpha, \beta_2) = \|\alpha - \beta\|.$$

Mostrar que

(a) $d(\alpha, \beta) \geq 0$;

(b) $d(\alpha, \beta) = 0$ se, e somente se, $\alpha = \beta$;

(c) $d(\alpha, \beta) = d(\beta, \alpha)$;

(d) $d(\alpha, \beta) \leq d(\alpha, \gamma) + d(\gamma, \beta)$.

5. Seja V um espaço com produto interno e sejam α, β vetores em V. Mostrar que $\alpha = \beta$ se, e somente se, $(\alpha|\gamma) = (\beta|\gamma)$ para todo γ em V.

6. Seja W o subespaço de R^2 gerado pelo vetor $(3, 4)$. Usando o produto interno canônico, seja E a projeção ortogonal de R^2 sobre W. Determinar:

(a) uma fórmula para $E(x_1, x_2)$;

(b) a matriz de E em relação à base ordenada canônica;

(c) $W^\perp$;

(d) uma base ortonormal em relação à qual E seja representada pela matriz

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

7. Seja V o espaço com produto interno que consiste de R^2 com o produto interno cuja forma quadrática é definida por

$$\|(x_1, x_2)\|^2 = (x_1 - x_2)^2 + 3x_2^2.$$

Seja E a projeção ortogonal de V sobre o subespaço W gerado pelo vetor $(3, 4)$. Responder agora às quatro questões do Exercício 6.

8. Determinar um produto interno sobre R^2 tal que $(\varepsilon_1, \varepsilon_2) = 2$.

9. Seja V o subespaço de $R[x]$ formado pelos polinômios de grau no máximo 3. Equipemos V com o produto interno

$$(f \mid g) = \int_0^1 f(t)\, g(t)\, dt.$$

(a) Determinar o suplementar ortogonal do subespaço dos polinômios constantes.

(b) Aplicar o processo de Gram-Schmidt à base $\{1, x, x^2, x^3\}$.

10. Seja V o espaço vetorial das $n \times n$ matrizes sobre C, com o produto interno $(A \mid B) = \operatorname{tr}(AB^*)$. Determinar o suplementar ortogonal do subespaço das matrizes diagonais.

11. Seja V um espaço de dimensão finita com produto interno e seja $\{\alpha_1, \ldots, \alpha_n\}$ uma base ortonormal de V. Mostrar que para quaisquer vetores α, β em V

$$(\alpha \mid \beta) = \sum_{k=1}^{n} (\alpha \mid \alpha_k)\, \overline{(\beta \mid \alpha_k)}.$$

12. Seja W um subespaço de dimensão finita de um espaço V com produto interno e seja E a projeção ortogonal de V sobre W. Demonstrar que $(E\alpha \mid \beta) = (\alpha \mid E\beta)$ para todos α, β em V.

13. Seja S um subconjunto de um espaço V com produto interno. Mostrar que $(S^\perp)^\perp$ contém o subespaço gerado por S. Para V de dimensão finita, mostrar que $(S^\perp)^\perp$ é o subespaço gerado por S.

14. Seja V um espaço de dimensão finita com produto interno e seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base ortonormal de V. Seja T um operador linear sobre V e A a matriz de T em relação à base ordenada $\mathscr{B}$. Demonstrar que

$$A_{ij} = (T\alpha_j \mid \alpha_i).$$

15. Suponhamos que $V = W_1 \oplus W_2$ e que f_1 e f_2 sejam produtos internos sobre W_1 e W_2 respectivamente. Mostrar que existe um único produto interno f sobre V tal que

(i) $W_2 = W_1^{\perp}$;

(ii) $f(\alpha, \beta) = f_k(\alpha, \beta)$, quando α, β estão em W_k, $k = 1, 2$.

16. Seja V um espaço com produto interno e W um subespaço de V de dimensão finita. Existem (em geral) muitas projeções que têm W por sua imagem. Uma destas, a projeção ortogonal sobre W, tem a propriedade de que $\|E\alpha\| \leq \|\alpha\|$ para todo α em V. Demonstrar que se E é uma projeção com imagem W, tal que $\|E\alpha\| \leq \|\alpha\|$ para todo α em V, então E é a projeção ortogonal sobre W.

17. Seja V o espaço real com produto interno que consiste do espaço das funções contínuas, definidas no intervalo $-1 \leq t \leq 1$, tomando valores reais, com o produto interno

$$(f \mid g) = \int_{-1}^{1} f(t)\, g(t)\, dt.$$

Seja W o subespaço das funções ímpares, isto é, funções que satisfazem $f(-t) = -f(t)$. Determinar o suplementar ortogonal de W.

8.3 Funcionais Lineares e Adjuntos

A primeira parte desta seção trata dos funcionais lineares sobre um espaço com produto interno e de sua relação com o produto interno. O resultado fundamental é que todo funcional linear f sobre um espaço de dimensão finita com produto interno é o "produto interno por um vetor fixo no espaço", isto é, que um tal f é da forma $f(\alpha) = (\alpha \mid \beta)$, para um certo β fixo em V. Usaremos este resultado para demonstrar a existência do "adjunto" de um operador linear T sobre V, sendo este um operador linear T^* tal que $(T\alpha \mid \beta) = (\alpha \mid T^*\beta)$ para todos α e β em V. Através do uso de uma base ortonormal, esta operação de conjugação sobre operadores lineares (passando de T a T^*) é identificada com a operação de se tomar a transposta conjugada de uma matriz. Vamos explorar superficialmente a analogia entre a operação de conjugação e a conjugação sobre números complexos.

Seja V um espaço arbitrário com produto interno e seja β um certo vetor fixo em V. Definamos uma função f_β de V no corpo de escalares por

$$f_\beta(\alpha) = (\alpha \mid \beta).$$

Esta função f_β é um funcional linear sobre V, pois, por sua própria definição, $(\alpha|\beta)$ é linear como uma função de α. Se V é a dimensão finita, todo funcional linear sobre V provém desta maneira de algum β.

Teorema 6. *Seja* **V** *um espaço de dimensão finita com produto interno e* f *um funcional linear sobre* **V**. *Então existe um único vetor* β *em* **V** *tal que* $\mathrm{f}(\alpha) = (\alpha|\beta)$ *para todo* α *em* **V**.

Demonstração. Seja $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ uma base ortonormal de V. Coloquemos

(8-13) $$\beta = \sum_{j=1}^{n} \overline{f(\alpha_j)}\alpha_j$$

e seja f_β o funcional linear definido por

$$f_{\beta(\alpha)} = (\alpha|\beta)$$

Então

$$f_\beta(\alpha_k) = (\alpha_k|\sum_j \overline{f(\alpha_j)}\alpha_j) = f(\alpha_k).$$

Como isto é válido para todo α_k, decorre que $f = f_\beta$. Suponhamos agora que γ seja um vetor em V tal que $(\alpha|\beta) = (\alpha|\gamma)$ para todo α. Então $(\beta - \gamma|\beta - \gamma) = 0$ e $\beta = \gamma$. Assim, existe exatamente um vetor β que determina o funcional linear f da maneira afirmada.

A demonstração deste teorema pode ser ligeiramente reformulada, em termos da representação de funcionais lineares em relação a uma base. Se tomarmos uma base ortonormal $\{\alpha_1, \ldots, \alpha_n\}$ de V, o produto interno de $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$ e $\beta = y_1\alpha_1 + \ldots + y_n\alpha_n$ será

$$(\alpha|\beta) = x_1 y_1 + \ldots + x_n y_n.$$

Se f é um funcional linear arbitrário sobre V, então f é da forma

$$f(\alpha) = c_1 x_1 + \ldots + c_n x_n$$

para certos escalares fixos $c_1, \ldots, c_n$ determinados pela base. É claro que $c_j = f(\alpha_j)$. Se desejamos encontrar um vetor β em V tal que $(\alpha|\beta) = \mathrm{f}(\alpha)$ para todo α, então evidentemente as coordenadas y_j de β devem satisfazer $y_j = c_j$, ou seja, $y_j = \overline{f(\alpha_j)}$. Conseqüentemente

$$\beta = \overline{f(\alpha_1)}\alpha_1 + \ldots + \overline{f(\alpha_n)}\alpha_n$$

é o vetor desejado.

Alguns comentários adicionais devem ser feitos. A demonstração do Teorema 6, que acabamos de fazer, é admiravelmente curta, mas ela deixa de dar ênfase ao fato geométrico essencial de que β está no suplementar ortogonal do núcleo de f. Seja W o núcleo de f. Então $V = W + W^{\perp}$, e f é completamente determinada pelos seus valores em $W^{\perp}$. De fato, se P é a projeção ortogonal de V sobre $W^{\perp}$, então

$$f(\alpha) = f(P\alpha)$$

para todo α em V. Suponhamos $f \neq 0$. Então, o posto de f é 1 e $\dim(W^{\perp}) = 1$. Se γ é um vetor não nulo, arbitrário, de $W^{\perp}$, segue-se que

$$P\alpha = \frac{(\alpha|\gamma)}{\|\gamma\|^2}\gamma$$

para todo α em V. Assim

$$f(\alpha) = (\alpha|\gamma) \cdot \frac{f(\gamma)}{\|\gamma\|^2}$$

para todo α, e $\beta = [\overline{f(\gamma)}/\|\gamma\|^2]\,\gamma$.

Exemplo 16. Gostaríamos de dar um exemplo que mostre que o Teorema 6 não é válido sem a hipótese de V ser de dimensão finita. Seja V o espaço vetorial dos polinômios sobre o corpo dos números complexos, com o produto interno

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\,dt.$$

Este produto interno pode também ser definido algebricamente. Se $f = \Sigma a_k x^k$ e $g = \Sigma b_k x^k$, então

$$(f|g) = \sum_{j,k} \frac{1}{j+k+1} a_j \overline{b_k}.$$

Seja z um número complexo fixo e seja L o funcional linear "valor que assume em z":

$$L(f) = f(z).$$

Existe um polinômio g tal que $(f|g) = L(f)$ para todo f? A resposta é negativa; de fato, suponhamos que se tenha

$$f(z) = \int_0^1 f(t)\overline{g(t)}\,dt$$

para todo f. Seja $h = x - z$, de modo que para todo f temos $(hf)(z) = 0$. Então

$$0 = \int_0^1 h(t)f(t)\overline{(gt)}\,dt$$

para todo f. Em particular, isto vale para $f = \bar{h}g$ de modo que

$$\int_0^1 |h(t)|^2|g(t)|^2\,dt = 0$$

e então $hg = 0$. Como $h \neq 0$, devemos ter que $g = 0$. Mas L não é o funcional nulo; logo, nenhum tal g existe.

Pode-se, num certo sentido, generalizar o exemplo, para o caso em que L é uma combinação linear de funcionais do tipo acima. Suponhamos que tomemos números complexos fixos $z_1, \ldots, z_n$ e escalares $c_1, \ldots, c_n$ e seja

$$L(f) = c_1 f(z_1) + \ldots + c_n f(z_n).$$

Então L é um funcional linear sobre V, mas não existe nenhum g tal que $L(f) = (f|g)$, a menos que $c_1 = c_2 = \ldots = c_n = 0$. Basta repetir o argumento acima com $h = (x - z_1) \ldots (x - z_n)$.

Voltamos agora ao conceito do adjunto de um operador linear.

Teorema 7. *Para qualquer operador linear* T *sobre um espaço* V *de dimensão finita com produto interno, existe um único operador linear* T* *sobre* V *tal que*

(8-12) $$(T\alpha|\beta) = (\alpha|T^*\beta)$$

para todos α, β *em* V.

Demonstração. Seja β um vetor arbitrário em V. Então $\alpha \to (T\alpha|\beta)$ é um funcional linear sobre V. Pelo Teorema 6, existe

um único vetor β' em V tal que $(T\alpha|\beta) = (\alpha|\beta')$ para todo α em V. Indiquemos por T^* a aplicação $\beta \to \beta'$:

$$\beta' = T^*\beta.$$

Temos (8-14), mas precisamos verificar que T^* é um operador linear. Sejam β, γ em V e seja c um escalar. Então, para qualquer α

$$\begin{aligned}(\alpha|T^*(c\beta + \gamma)) &= (T\alpha|c\beta + \gamma)\\ &= (T\alpha|c\beta) + (T\alpha|\gamma)\\ &= \bar{c}(T\alpha|\beta) + (T\alpha|\gamma)\\ &= \bar{c}(\alpha|T^*\beta) + (\alpha|T^*\gamma)\\ &= (\alpha|cT^*\beta) + (\alpha|T^*\gamma)\\ &= (\alpha|cT^*\beta + T^*\gamma).\end{aligned}$$

Assim, $T^*(c\beta + \gamma) = cT^*\beta + T^*\gamma$ e T^* é linear.

A unicidade de T^* é evidente. Para β arbitrário em V, o vetor $T^*\beta$ é determinado de modo único como sendo o vetor β' tal que $(T\alpha|\beta) = (\alpha|\beta')$ para todo α.

Teorema 8. *Seja* V *um espaço de dimensão finita com produto interno e seja* $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ *uma base ortonormal (ordenada) de* V. *Seja* T *um operador linear sobre* V *e seja* A *a matriz de* T *em relação à base ordenada* $\mathscr{B}$. *Então* $A_{kj} = (T\alpha_j|\alpha_k)$.

Demonstração. Como $\mathscr{B}$ é uma base ortonormal, temos

$$\alpha = \sum_{k=1}^{n} (\alpha|\alpha_k)\,\alpha_k.$$

A matriz A é definida por

$$T\alpha_j = \sum_{k=1}^{n} A_{kj}\alpha_k$$

e como

$$T\alpha_j = \sum_{k=1}^{n} (T\alpha_j|\alpha_k)\,\alpha_k$$

temos $A_{kj} = (T\alpha_j|\alpha_k)$.

Corolário. *Seja* V *um espaço de dimensão finita com produto interno e seja* T *um operador linear sobre* V. *Em relação a qualquer*

base ortonormal de V, *a matriz de* T* *é a transposta conjugada da matriz de* T.

Demonstração. Seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base ortonormal de V, seja $A = [T]_{\mathscr{B}}$ e $B = [T^*]_{\mathscr{B}}$. De acordo com o Teorema 8.

$$\begin{aligned} A_{kj} &= (T\alpha_j|\alpha_k) \\ B_{kj} &= (T^*\alpha_j|\alpha_k). \end{aligned}$$

Então, pela definição de T^*, temos

$$\begin{aligned} B_{kj} &= (T^*\alpha_j|\alpha_k) \\ &= \overline{(\alpha_k|T^*\alpha_j)} \\ &= \overline{(T^*_k|\alpha_j)} \\ &= \overline{A}_{jk}. \end{aligned}$$

Exemplo 17. Seja V um espaço de dimensão finita com produto interno e seja E a projeção ortogonal de V sobre um subespaço W. Então, para quaisquer vetores α e β em V,

$$\begin{aligned} (E\alpha|\beta) &= (E\alpha|E\beta + (1 - E)\beta) \\ &= (E\alpha|E\beta) \\ &= (E\alpha + (1 - E)\alpha|E\beta) \\ &= (\alpha|E\beta). \end{aligned}$$

Da unicidade do operador E^* segue-se que $E^* = E$. Consideremos agora a projeção E descrita no Exemplo 14. Então

$$A = \frac{1}{154}\begin{bmatrix} 9 & 36 & -3 \\ 36 & 144 & -12 \\ -3 & -12 & 1 \end{bmatrix}$$

é a matriz de E em relação à base ortonormal canônica. Como $E = E^*$, A é também a matriz de E^* e, porque $A = A^*$, isso não contradiz o corolário anterior. Por outro lado, suponhamos

$$\begin{aligned} \alpha_1 &= (154, 0, 0) \\ \alpha_2 &= (145, -36, 3) \\ \alpha_3 &= (-36, 10, 12). \end{aligned}$$

Então $\{\alpha_1, \alpha_2, \alpha_3\}$ é uma base e

$$\begin{aligned} E\alpha_1 &= (9, 36, -3) \\ E\alpha_2 &= (0, 0, 0) \\ E\alpha_3 &= (0, 0, 0). \end{aligned}$$

Como $(9, 36, -3) = -(154, 0, 0) - (145, -36, 3)$, a matriz B de E em relação à base $\{\alpha_1, \alpha_2, \alpha_3\}$ é definida pela equação

$$B = \begin{bmatrix} -1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Nesse caso $B \neq B^*$ e B^* não é a matriz de $E^* = E$ em relação à base $\{\alpha_1, \alpha_2, \alpha_3\}$. Aplicando o corolário, concluímos que $\{\alpha_1, \alpha_2, \alpha_3\}$ não é uma base ortonormal. De qualquer modo, tudo isso é trivial.

Definição. *Seja* T *um operador linear sobre um espaço* V *com produto interno. Dizemos que* T **possui um adjunto** *sobre* V *se existe um operador linear* T* *sobre* V *tal que* $(T\alpha|\beta) = (\alpha|T^*\beta)$ *para todos* α *e* β *em* V.

O Teorema 7 afirma que todo operador linear sobre um espaço V de dimensão finita com produto interno possui um adjunto sobre V. No caso de dimensão infinita isso nem sempre é verdade. Mas, em qualquer caso, existe no máximo um tal operador T^*; quando existir, denominamo-lo o **adjunto** de T.

Dois comentários devem ser feitos acerca do caso de dimensão finita.

1. O adjunto de T depende não só de T, mas também do produto interno.

2. Como foi visto no Exemplo 17, para uma base ordenada arbitrária $\mathscr{B}$, a relação entre $[T]_{\mathscr{B}}$ e $[T^*]_{\mathscr{B}}$ é mais complicada que a apresentada no corolário acima.

Exemplo 18. Seja $V = C^{n \times 1}$ o espaço das $n \times 1$ matrizes complexas, com o produto interno $(X|Y) = Y^*X$. Se A é uma $n \times n$ matriz com elementos complexos, o adjunto do operador linear $X \to AX$ é o operador $X \to A^*X$. De fato, pois

$$(AX|Y) = Y^*AX = (A^*Y)^*X = (X|A^*Y).$$

O leitor deverá se convencer de que isto é na realidade uma reformulação do último corolário.

Exemplo 19. Este é semelhante ao Exemplo 18. Seja $V = C^{n \times n}$ com produto interno $(A|B) = \operatorname{tr}(B^*A)$. Seja M uma $n \times n$ matriz fixa sobre C. O adjunto da multiplicação à esquerda por M é a multiplicação à esquerda por M^*. Evidentemente, "multiplicação à esquerda por M" é o operador linear L_M definido por $L_M(A) = MA$.

$$\begin{aligned}(L_M(A)|B) &= \operatorname{tr}(B^*(MA))\\ &= \operatorname{tr}(MAB^*)\\ &= \operatorname{tr}(AB^*M)\\ &= \operatorname{tr}(A(M^*B)^*)\\ &= (A|L_M{}^*(B)).\end{aligned}$$

Assim, $(L_M)^* = L_{M^*}$. No cálculo acima, usamos duas vezes a propriedade característica da função traço: $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Exemplo 20. Seja V o espaço dos polinômios sobre o corpo dos números complexos, com o produto interno

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\,dt.$$

Se f é um polinômio, $f = \Sigma\, a_k x^k$, seja $\bar{f} = \Sigma\, \bar{a}_k x^k$. Isto é, $\bar{f}$ é o polinômio cuja função polinomial associada é a complexa conjugada da de f:

$$\bar{f}(t) = \overline{f(t)}, \quad t \text{ real}$$

Consideremos o operador "multiplicação por f", isto é, o operador linear M_f definido por $M_f(g) = fg$. Então este operador possui um adjunto, a saber, a multiplicação por $\bar{f}$. De fato, pois

$$\begin{aligned}(M_f(g)|h) &= (fg|h)\\ &= \int_0^1 f(t)g(t)\overline{h(t)}\,dt\\ &= \int_0^1 g(t)[\overline{\bar{f}(t)h(t)}]\,dt\end{aligned}$$

$$= (g|\bar{f}h)$$
$$= (g|M_{\bar{f}}(h))$$

e, portanto, $(M_{\bar{f}})^* = M_f$.

Exemplo 21. No Exemplo 20, vimos que alguns operadores lineares sobre um espaço de dimensão infinita com produto interno possuem um adjunto. Como comentamos anteriormente, outros não o têm. Seja V o espaço com produto interno do Exemplo 20 e seja D o operador derivação sobre $C[x]$. A integração por partes mostra que

$$(Df|g) = f(1)g(1) - f(0)g(0) - (f|Dg).$$

Fixemos g e perguntemos quando é que existe um polinômio D^*g tal que $(Df|g) = (f|D^*g)$ para todo f. Se um tal D^*g existe, temos

$$(f|D^*g) = f(1)g(1) - f(0)g(0) - (f|Dg)$$

ou seja,

$$(f|D^*g + Dg) = f(1)g(1) - f(0)g(0).$$

Com g *fixo*, $L(f) = f(1)g(1) - f(0)g(0)$ é um funcional linear do tipo considerado no Exemplo 16 e não pode ser da forma $L(f) = (f|h)$ a menos que $L = 0$. Se D^*g existe, então com $h = D^*g + Dg$ temos de fato $L(f) = (f|h)$, e então $g(0) = g(1) = 0$. A existência de um polinômio adequado D^*g implica $g(0) = g(1) = 0$. Reciprocamente, se $g(0) = g(1) = 0$, o polinômio $D^*g = -Dg$ satisfaz $(Df|g) = (f|D^*g)$ para todo f. Tomando um g qualquer para o qual $g(0) \neq 0$ ou $g(1) \neq 0$, não podemos definir D^*g de modo conveniente; portanto, concluímos que D não possui adjunto.

Esperamos que estes exemplos aumentem a compreensão do leitor quanto ao adjunto de um operador linear. Vemos que a operação de conjugação, que faz passar de T a T^*, se comporta um pouco como a conjugação sobre números complexos. O teorema seguinte fortalece esta analogia.

Teorema 9. *Seja* V *um espaço de dimensão finita com produto interno. Se* T *e* U *são operadores lineares sobre* V *e* c *é um escalar*

(i) $(T + U)^* = T^* + U^*$;

(ii) $(cT)^* = \bar{c}T^*$;

(iii) $(TU)^* = U^*T^*$;
(iv) $(T^*)^* = T$.

Demonstração. Para demonstrar (i)

$$\begin{aligned}((T + U)\alpha|\beta) &= (T\alpha + U\alpha|\beta)\\ &= (T\alpha|\beta) + (U\alpha|\beta)\\ &= (\alpha|T^*\beta) + (\alpha|U^*\beta)\\ &= (\alpha|T^*\beta + U^*\beta)\\ &= (\alpha|(T^* + U^*)\beta).\end{aligned}$$

Pela unicidade do adjunto, temos $(T + U)^* = T^* + U^*$. Deixamos a demonstração de (ii) a cargo do leitor. Obtemos (iii) e (iv) a partir de

$$(TU\alpha|\beta) = (U\alpha|T^*\beta) = (\alpha|U^*T^*\beta)$$
$$(T^*\alpha|\beta) = \overline{(\beta|T^*\alpha)} = \overline{(T\beta|\alpha)} = (\alpha|T\beta).$$

O Teorema 9 é freqüentemente formulado como segue: a aplicação $T \rightarrow T^*$ é um antiisomorfismo linear-conjugado de período 2. A analogia com a conjugação complexa que mencionamos acima é, evidentemente, baseada na observação de que a conjugação complexa tem as propriedades $\overline{(z_1 + z_2)} = \bar{z}_1 + \bar{z}_2$, $\overline{(z_1 z_2)} = \bar{z}_1\bar{z}_2$, $\bar{\bar{z}} = z$. Deve-se ter o cuidado de observar a inversão da ordem num produto, imposta pela operação de conjugação: $(UT)^* = T^*U^*$. Mencionaremos extensões da analogia à medida que prossigamos nosso estudo de operadores lineares sobre um espaço com produto interno. Podemos mencionar alguma coisa nesse sentido agora. Um número complexo z é real se, e somente se, $z = \bar{z}$. É de esperar que os operadores lineares T tais que $T = T^*$ se comportem, de certa maneira, como os números reais. É isto o que realmente ocorre. Por exemplo, se T é um operador linear sobre um espaço *complexo* de dimensão finita com produto interno, então

(8-15) $$T = U_1 + iU_2$$

onde $U_1 = U^*_1$ e $U_2 = U^*_2$. Assim, de certa forma, T possui uma "parte real" e uma "parte imaginária". Os operadores U_1 e U_2 que satisfazem $U_1 = U^*_1$, $U_2 = U^*_2$ e (8-15) são únicos e são dados por

$$U_1 = \frac{1}{2}(T + T^*)$$

$$U_2 = \frac{1}{2i}(T - T^*).$$

Um operador linear T tal que $T = T^*$ é dito **auto-adjunto** (ou **hermitiano**). Se $\mathscr{B}$ é uma base ortonormal de V, então

$$[T^*]_{\mathscr{B}} = [T]^*_{\mathscr{B}}$$

e, então, T é auto-adjunto se, e somente se, sua matriz em relação a toda base ortonormal é uma matriz auto-adjunta. Operadores auto-adjuntos são importantes, não só porque nos fornecem uma espécie de partes real e imaginária de um operador linear arbitrário, mas também pelas seguintes razões: (1) Operadores auto-adjuntos possuem muitas propriedades especiais. Por exemplo, para um operador deste tipo, existe uma base ortonormal formada por vetores característicos. (2) Muitos operadores que surgem na prática são auto-adjuntos. Consideraremos posteriormente as propriedades especiais dos operadores auto-adjuntos.

Exercícios

1. Seja V o espaço C^2, com o produto interno canônico. Seja T o operador definido por $T\varepsilon_1 = (1, -2)$, $T\varepsilon_2 = (i, -1)$. Se $\alpha = (x_1^2, x_2)$, determinar $T^*\alpha$.

2. Seja T o operador linear sobre C^2 definido por $T\varepsilon_1 = (1 + i, 2)$, $T\varepsilon_2 = (i, i)$. Usando o produto interno canônico, determinar a matriz de T^* em relação à base ordenada canônica. T comuta com T^*?

3. Suponhamos que V seja C^3 com o produto interno canônico. Seja T o operador linear sobre V cuja matriz em relação à base ordenada canônica é definida por

$$A_{jk} = i^{j+k}, \quad (i^2 = -1).$$

Determinar uma base do núcleo de T^*.

4. Seja V um espaço de dimensão finita com produto interno e T um operador linear sobre V. Mostrar que a imagem de T^* é o suplementar ortogonal do núcleo de T.

5. Seja V um espaço de dimensão finita com produto interno e T um operador linear sobre V. Se T é inversível, mostrar que T^* é inversível e $(T^*)^{-1} = (T^{-1})^*$.

6. Seja V um espaço com produto interno e β, γ vetores fixos em V. Mostrar que $T\alpha = (\alpha|\beta)\gamma$ define um operador linear sobre V. Mostrar que T possui um adjunto e descrever T^* explicitamente.

Suponhamos agora que V seja C^n com o produto interno canônico, $\beta = (y_1, \ldots, y_n)$ e $\gamma = (x_1, \ldots, x_n)$. Qual é o elemento j, k da matriz de T em relação à base ordenada canônica? Qual é o posto desta matriz?

7. Mostrar que o produto de dois operadores auto-adjuntos é auto-adjunto se, e somente se, os dois operadores comutam.

8. Seja V o espaço vetorial dos polinômios sobre R de grau menor ou igual a 3, com o produto interno

$$(f|g) = \int_0^1 f(t)\, g(t)\, dt.$$

Se t é um número real, determinar o polinômio g_t em V tal que $(f|g_t) = f(t)$ para todo f em V.

9. Seja V o espaço com produto interno do Exercício 8 e seja D operador derivação sobre V. Determinar D^*.

10. Seja V o espaço das $n \times n$ matrizes sobre o corpo dos números complexos, com o produto interno $(A|B) = \operatorname{tr}(AB^*)$. Seja P uma matriz inversível fixa em V e seja T_P o operador linear sobre V definido por $T_P(A) = P^{-1}AP$. Determinar o adjunto de T_P.

11. Seja V um espaço de dimensão finita com produto interno e seja E um operador linear idempotente sobre V, isto é, $E^2 = E$. Demonstrar que E é auto-adjunto se, e somente se, $EE^* = E^*E$.

12. Seja V um espaço *complexo* de dimensão finita com produto interno e seja T um operador linear sobre V. Demonstrar que T é auto-adjunto se, e somente se, $(T\alpha|\alpha)$ é real para todo α em V.

8.4 Operadores Unitários

Nesta seção, vamos considerar o conceito de um isomorfismo entre dois espaços com produto interno. Se V e W são espaços vetoriais, um isomorfismo de V em W é uma transformação linear bijetora de V em W, isto é, uma correspondência bijetora entre os elementos de V e os de W, a qual "conserva" as operações de espaço vetorial. Ora, um espaço com produto interno consiste de um espaço vetorial e um produto interno especificado sobre aquele espaço. Assim, quando V e W são espaços com produto interno, exigiremos que um isomorfismo de V em W não só conserve as operações lineares, mas também conserve produtos internos. Um isomorfismo de um espaço com produto interno em si mesmo é deno-

minado um "operador unitário" sobre aquele espaço. Consideraremos vários exemplos de operadores unitários e estabeleceremos suas propriedades fundamentais.

Definição. *Sejam* V *e* W *espaços com produto interno sobre o mesmo corpo e seja* T *uma transformação linear de* V *em* W. *Dizemos que* T **conserva produtos internos** *se* $(T\alpha|T\beta) = (\alpha|\beta)$ *para todos* α, β *em* V. *Um* **isomorfismo** *de* V *em* W *é um* **isomorfismo** T *de espaço vetorial de* **V** *em* **W** *que também conserva produtos internos.*

Se T conserva produtos internos, então $\|T\alpha\| = \|\alpha\|$, portanto, T é, necessariamente, não-singular. Assim, um isomorfismo de V em W pode também ser definido como uma transformação linear de V sobre W que conserva produtos internos. Se T é um isomorfismo de V em W, então T^{-1} é um isomorfismo de W em V; logo, quando um tal T existir, diremos simplesmente que V e W são **isomorfos**. É claro que o isomorfismo de espaços com produto interno é uma relação de equivalência.

Teorema 10. *Sejam* V *e* W *espaços de dimensão finita com produto interno sobre o mesmo corpo, que tenham a mesma dimensão. Se* T *é uma transformação linear de* V *em* W, *as seguintes afirmações são equivalentes:*

(i) T *conserva produtos internos.*

(ii) T *é um isomorfismo (de espaço com produto interno).*

(iii) T *leva toda base ortonormal de* V *em uma base ortonormal de* W.

(iv) T *leva alguma base ortonormal de* **V** *em alguma base ortonormal de* W.

Demonstração (i)$\rightarrow$(ii). Se T conserva produtos internos, então $\|T\alpha\| = \|\alpha\|$ para todo α em V. Assim, T é não-singular e como $\dim V = \dim W$, sabemos que T é um isomorfismo de espaço vetorial.

(ii)$\rightarrow$(iii) Suponhamos que T seja um isomorfismo. Seja $\{\alpha_1, \ldots, \alpha_n\}$ uma base ortonormal de V. Como T é um isomorfismo de espaço vetorial e $\dim W = \dim V$, decorre que $\{T\alpha_1, \ldots, T\alpha_n\}$ é uma base de W. Como T conserva também produtos internos, $(T\alpha_j|T\alpha_k) = (\alpha_j|\alpha_k) = \delta_{jk}$.

(iii) → (iv) Não requer comentários.

(iv) → (i) Seja $\{\alpha_1, \ldots, \alpha_n\}$ uma base ortonormal de V tal que $\{T\alpha_1, \ldots, T\alpha_n\}$ seja uma base ortonormal de W. Então

$$(T\alpha_j | T\alpha_k) = (\alpha_j | \alpha_k) = b_{jk}.$$

Para todos $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$ e $\beta = y_1\alpha_1 + \ldots + y_n\alpha_n$ em V, temos

$$(\alpha|\beta) = \sum_{j=1}^{n} x_j\bar{y}_j$$

$$\begin{aligned}(T\alpha|T\beta) &= (\sum_j x_j T\alpha_j | \sum_k y_k T\alpha_k) \\ &= \sum_j \sum_k x_j \bar{y}_k (T\alpha_j | T\alpha_k) \\ &= \sum_{j=1}^{n} x_j \bar{y}_j\end{aligned}$$

logo T conserva produtos internos.

Corolário. *Sejam* V *e* W *espaços de dimensão finita como produto interno sobre o mesmo corpo. Então* V *e* W *são isomorfos se, e somente se, têm a mesma dimensão.*

Demonstração. Se $\{\alpha_1, \ldots, \alpha_n\}$ é uma base ortonormal de V e $\{\beta_1, \ldots, \beta_n\}$ é uma base ortonormal de W, seja T a transformação linear de V em W definida por $T\alpha_j = \beta_j$. Então T é um isomorfismo de V em W.

Exemplo 22. Se V é um espaço n-dimensional com produto interno, então toda base ortonormal ordenada $\mathscr{B} = \{\alpha, \ldots, \alpha_n\}$ determina um isomorfismo de V em F^n com o produto interno canônico. O isomorfismo é simplesmente

$$T(x_1\alpha_1 + \ldots + x_n\alpha_n) = (x_1, \ldots, x_n).$$

Existe o isomorfismo, superficialmente diferente, determinado por $\mathscr{B}$, de V no espaço $F^{n\times 1}$ com $(X|Y) = Y^*X$ como produto interno. O isomorfismo é

$$\alpha \to [\alpha]_{\mathscr{B}}$$

isto é, a transformação que leva α na matriz de suas coordenadas em relação à base ordenada $\mathscr{B}$. Para qualquer base ordenada $\mathscr{B}$, este é um isomorfismo de espaço vetorial; no entanto, ele é um isomorfismo dos dois espaços com produto interno se, e somente se, $\mathscr{B}$ é ortonormal.

Exemplo 23. Eis um isomorfismo um pouco menos superficial. Seja W o espaço das 3×3 matrizes A sobre R que sejam anti-simétricas, isto é, $A^t = -A$. Vamos equipar W com o produto interno $(A|B) = \frac{1}{2}\operatorname{tr}(AB^t)$, sendo o $\frac{1}{2}$ colocado por conveniência. Seja V o espaço R^3 com o produto interno canônico. Seja T a transformação linear de V em W definida por

$$T(x_1, x_2, x_3) = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}.$$

Então T leva V sobre W, e colocando

$$A = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix}$$

temos

$$\begin{aligned} \operatorname{tr}(AB^t) &= x_3y_3 + x_2y_2 + x_3y_3 + x_1y_1 + x_2y_2 + x_1y_1 \\ &= 2(x_1y_1 + x_2y_2 + x_3y_3). \end{aligned}$$

Assim, $(\alpha|\beta) = (T\alpha|T\beta)$ e T é um isomorfismo de espaço vetorial. Notemos que T leva a base canônica $\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ na base ortonormal formada pelas três matrizes

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Exemplo 24. Nem sempre é particularmente conveniente descrever um isomorfismo em termos de bases ortonormais. Por exemplo, suponhamos $G = P^*P$, onde P é uma $n \times n$ matriz inversível com elementos complexos. Seja V o espaço das $n \times 1$ matrizes complexas, com o produto interno $[X|Y] = Y^*GX$. Seja W o mesmo espaço vetorial, com o produto interno canônico $(X|Y) = Y^*X$. Sabemos que V e W são espaços com produto

interno que são isomorfos. Deveria parecer que a maneira mais conveniente de descrever um isomorfismo entre V e W seja a seguinte: Seja T a transformação linear de V em W definida por $T(X) = PX$. Então

$$\begin{aligned}(TX|TY) &= (PX|PY)\\ &= (PY)^*(PX)\\ &= Y^*P^*PX\\ &= Y^*GX\\ &= [X|Y].\end{aligned}$$

Logo, T é um isomorfismo.

Exemplo 25. Seja V o espaço das funções contínuas, definidas sobre o intervalo unitário, $0 \leq t \leq 1$ e tomando valores reais, com o produto interno

$$[(f|g)] = \int_0^1 f(t)g(t)t^2 dt.$$

Seja W o mesmo espaço vetorial com o produto interno

$$(f|g) = \int_0^1 f(t)g(t)\,dt.$$

Seja T a transformação linear de V em W dada por

$$(Tf)(t) = tf(t).$$

Então $(Tf|Tg) = [f|g]$, portanto T conserva produtos internos; contudo, T *não* é um isomorfismo de V em W, porque a imagem de T não é todo o espaço W. Evidentemente, isto ocorre porque o espaço vetorial subjacente não é de dimensão finita.

Teorema 11. *Sejam* V *e* W *espaços com produto interno, sobre o mesmo corpo e seja* T *uma transformação linear de* V *em* W. *Então,* T *conserva produtos internos se, e somente se,* $||T\alpha|| = ||\alpha||$ *para todo* α *em* V.

Demonstração. Se T conserva produtos internos, T "conserva normas". Suponhamos que $||T\alpha|| = ||\alpha||$ para todo α em V. Então $||T\alpha||^2 = ||\alpha||^2$. Usando agora a identidade de polarização conveniente, (8-3) ou (8-4) e o fato de que T é linear, obtém-se facilmente $(\alpha|\beta = (T\alpha|T\beta)$ para quaisquer α, β em V.

Definição. *Um* **operador unitário** *sobre um espaço com produto interno é um isomorfismo do espaço em si mesmo.*

O produto de dois operadores unitários é unitário. De fato, se U_1 e U_2 são unitários, então U_2U_1 é inversível e $\|U_2U_1\alpha\| = \|U_1\alpha\| = \|\alpha\|$ para todo α. Além disso, o inverso de um operador unitário é unitário, pois $\|U\alpha\| = \|\alpha\|$ diz que $\|U^{-1}\beta\| = \|\beta\|$, onde $\beta = U\alpha$. Como o operador idêntico é obviamente unitário, vemos que o conjunto dos operadores unitários sobre um espaço com produto interno é um grupo, com a operação de composição.

Se V é um espaço de dimensão finita com produto interno e U é um operador linear sobre V, o Teorema 10 nos diz que U é unitário se, e somente se, $(U\alpha|U\beta) = (\alpha|\beta)$ para todos α, β em V; ou seja, se, e somente se, para alguma (toda) base ortonormal $\{\alpha_1, \ldots, \alpha_n\}$ é verdade que $\{U\alpha_1, \ldots, U\alpha_n\}$ é uma base ortonormal.

Teorema 12. *Seja* U *um operador linear sobre um espaço* V *com produto interno. Então* U *é unitário se, e somente se, o adjunto* U* *de* U *existe e* $UU^* = U^*U = I$.

Demonstração. Suponhamos que U seja unitário. Então U é inversível e

$$(U\alpha|\beta) = (U\alpha|UU^{-1}\beta) = (\alpha|U^{-1}\beta)$$

para todos α, β. Logo, U^{-1} é o adjunto de U.

Reciprocamente, suponhamos que U^* exista e $U^{**} = U^*U = I$. Então U é inversível, com $U^{-1} = U^*$. Portanto, basta mostrar que U conserva produtos internos. Temos

$$\begin{aligned}(U\alpha|U\beta) &= (\alpha|U^*U\beta)\\ &= (\alpha|I\beta)\\ &= (\alpha|\beta)\end{aligned}$$

para todos α, β.

Exemplo 26. Consideremos $C^{n\times 1}$, com o produto interno $(X|Y) = Y^*X$. Seja A uma $n \times n$ matriz sobre C e seja U o operador linear definido por $U(X) = AX$. Então

$$(UX|UY) = (AX|AY) = Y^*A^*AX$$

para todas X, Y. Logo, U é unitário se, e somente se, $A^*A = I$.

Definição. *Uma* n × n *matriz complexa* A *é dita* **unitária** *se* $A^*A = I$.

Teorema 13. *Seja* V *um espaço de dimensão finita com produto interno e seja* U *um operador linear sobre* V. *Então* U *é unitário se, e somente se, a matriz de* U *em relação a alguma (toda) base ortonormal ordenada é uma matriz unitária.*

Demonstração. A esta altura, isto não é bem um teorema e só o enunciamos por questão de ênfase. Se $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ é uma base ortonormal ordenada de V e A é a matriz de U em relação a $\mathscr{B}$, então $A^*A = I$ se, e somente se, $U^*U = I$. O resultado decorre do Teorema 12.

Seja A uma $n \times n$ matriz. A afirmação de que A é unitária diz simplesmente

$$(A^*A)_{jk} = \delta_{jk}$$

ou

$$\sum_{r=1}^{n} \bar{A}_{rj}A_{rk} = \delta_{jk}.$$

Em outras palavras, diz que as colunas de A formam um conjunto ortonormal de matrizes-colunas, usando o produto interno canônico $(X|Y) = Y^*X$. Como $A^*A = I$ se, e somente se, $AA^* = I$ vemos que A é unitária exatamente quando as linhas de A constituem um conjunto ortonormal de n-uplas em C^n (com o produto interno canônico). Portanto, usando produtos internos canônicos, A é unitária se, e somente se, as linhas e colunas de A são conjuntos ortonormais. Vê-se aqui um exemplo da força do teorema que afirma que uma inversa unilateral de uma matriz é, na verdade, uma inversa bilateral. Aplicando este teorema como fizemos acima, digamos, a matrizes reais, temos o seguinte: Suponhamos que exista uma tabela quadrada de números reais tal que a soma dos quadrados dos elementos de cada linha seja 1 e tal que linhas distintas sejam ortogonais. Então a soma dos quadrados dos elementos de cada coluna é 1 e colunas distintas são ortogonais. Basta escrever a demonstração deste fato para uma tabela 3 × 3, sem usar nenhum conhecimento sobre matrizes, para se ficar razoavelmente impressionado.

Definição. *Uma* n × n *matriz* A, *real ou complexa, é dita* **ortogonal** *se* $A^tA = I$.

Uma matriz ortogonal real é unitária; uma matriz unitária é ortogonal se, e somente se, cada um dos seus elementos é real.

Exemplo 27. Vejamos alguns exemplos de matrizes unitárias e ortogonais.

(a) Uma 1 × 1 matriz $[c]$ é ortogonal se, e somente se, $c = \pm 1$ e é unitária se, e somente se, $\bar{c}c = 1$. A última condição significa (é claro) que $|c| = 1$, ou seja, $c = e^{i\theta}$, com θ real.

(b) Seja

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Então A é ortogonal se, e somente se,

$$A^t = A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

O determinante de qualquer matriz ortogonal, como se pode ver facilmente, é ± 1. Assim, A é ortogonal se, e somente se,

$$A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

ou

$$A = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$$

onde $a^2 + b^2 = 1$. Os dois casos são distinguidos pelo valor de det A.

(c) As bem conhecidas relações entre as funções trigonométricas mostram que a matriz

$$A_\theta = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$$

é ortogonal. Se θ é um número real, então A_θ é a matriz, em relação à base ordenada canônica de R^2, do operador linear U_θ, que é a rotação do ângulo θ. A afirmação de que A_θ é uma matriz

ortogonal real (logo unitária) significa simplesmente que U_θ é um operador unitário, isto é, conserva o produto escalar.

(d) Seja

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Então A é unitária se, e somente se,

$$\begin{bmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

O determinante de uma matriz unitária tem valor absoluto 1 e é, portanto, um número complexo da forma $e^{i\theta}$, θ real. Assim, A é unitária se, e somente se,

$$A = \begin{bmatrix} a & b \\ -e^{i\theta}\bar{b} & e^{i\theta}\bar{a} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix} \begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix}$$

onde θ é um número real e a, b são números complexos tais que $|a|^2 + |b|^2 = 1$.

Como observamos anteriormente, os operadores unitários sobre espaços com produto interno formam um grupo. Desse fato e do Teorema 13 segue-se que o conjunto $U(n)$ das $n \times n$ matrizes unitárias também é um grupo. Assim, a inversa de uma matriz unitária e o produto de duas matrizes unitárias é novamente uma matriz unitária. É claro que é fácil verificar esses fatos diretamente. Uma $n \times n$ matriz A com elementos complexos é unitária se, e somente se, $A^{-1} = A^*$. Assim, se A é unitária, temos $(A^{-1})^{-1} = A = (A^*)^{-1} = (A^{-1})^*$. Se A e B são $n \times n$ matrizes unitárias, então $(AB)^{-1} = B^{-1}A^{-1} = B^*A^* = (AB)^*$.

O processo de Gram-Schmidt em C^n tem um corolário interessante para matrizes, que envolve o grupo $U(n)$.

Teorema 14. *Para cada* n × n *matriz complexa inversível* B, *existe uma única matriz triangular inferior* M, *com elementos positivos na diagonal principal, tal que* MB *é unitária.*

Demonstração. As linhas $\beta_1, \ldots, \beta_n$ de B formam uma base de C^n. Sejam $\alpha_1, \ldots, \alpha_n$ vetores obtidos de $\beta_1, \ldots, \beta_n$ pelo processo de Gram-Schmidt. Então, para $1 \leq k \leq n$, $\{\alpha_1, \ldots, \alpha_k\}$ é uma base ortogonal do subespaço gerado por $\{\beta_1, \ldots, \beta_k\}$ e

$$\alpha_k = \beta_k - \sum_{j<k} \frac{(\beta_k|\alpha_j)}{\|\alpha_j\|^2} \alpha_j.$$

Logo, para cada k existem escalares C_{kj}, determinados de modo único, tais que

$$\alpha_k = \beta_k - \sum_{j<k} C_{kj}\beta_j.$$

Seja U a matriz unitária com linhas

$$\frac{\alpha_1}{\|\alpha_1\|}, \ldots, \frac{\alpha_n}{\|\alpha_n\|}$$

e M a matriz definida por

$$M_{kj} = \begin{cases} -\dfrac{1}{\|\alpha_k\|} \cdot C_{kj}, & \text{se } j < k \\ \dfrac{1}{\|\alpha_k\|}, & \text{se } j = k \\ 0, & \text{se } j > k. \end{cases}$$

Então M é triangular inferior, no sentido de que os elementos acima da diagonal principal são nulos. Os elementos M_{kk} de M na diagonal principal são todos positivos, e

$$\frac{\alpha_k}{\|\alpha_k\|} = \sum_{j=1}^{n} M_{kj}\beta_j, \qquad 1 \leq k \leq n.$$

Ora, essas equações simplesmente dizem que

$$U = MB.$$

Para demonstrar a unicidade de M, indiquemos por $T^+(n)$ o conjunto de todas as $n \times n$ matrizes complexas triangular inferiores com elementos positivos na diagonal principal. Suponhamos que M_1 e M_2 sejam elementos de $T^+(n)$ tais que M_iB esteja em $U(n)$ para $i = 1, 2$. Como $U(n)$ é um grupo,

$$(M_1B)(M_2B)^{-1} = M_1M_2^{-1}$$

está em $U(n)$. Por outro lado, apesar de não ser inteiramente óbvio, $T^+(n)$ também é um grupo em relação à multiplicação de matrizes. Uma das maneiras de se ver isso é a de considerar as propriedades geométricas das transformações lineares

$$X \to MX, \qquad (M \text{ em } T^+(n))$$

sobre o espaço das matrizes-coluna. Assim, M_2^{-1}, $M_1M_2^{-1}$ e $(M_1M_2^{-1})^{-1}$ estão todas em $T^+(n)$. Mas, como $M_1M_2^{-1}$ está em $U(n)$, $(M_1M_2^{-1})^{-1} = (M_1M_2^{-1})^*$. A transposta ou a transposta conjugada de qualquer matriz triangular inferior é uma matriz triangular superior. Portanto, $M_1M_2^{-1}$ é simultaneamente triangular superior e inferior, isto é, diagonal. Uma matriz diagonal é unitária se, e somente se, cada um de seus elementos na diagonal principal tiver valor absoluto 1; se os elementos diagonais forem todos positivos, eles deverão ser iguais a 1. Logo, $M_1M_2^{-1} = I$ e $M_1 = M_2$.

Indiquemos por $GL(n)$ o conjunto das $n \times n$ matrizes complexas inversíveis. Então $GL(n)$ é também um grupo em relação à multiplicação de matrizes. Esse grupo é denominado o **grupo linear geral.** O Teorema 14 é equivalente ao seguinte resultado:

Corolário. *Para cada* B *em* GL(n) *existem matrizes* N *e* U, *determinadas de modo único, tais que* N *está em* $T^+(n)$, U *em* U(n) *e*

$$B = N \cdot U.$$

Demonstração. Pelo teorema, existe uma única matriz M em $T^+(n)$ tal que MB esteja em $U(n)$. Seja $MB = U$ e $N = M^{-1}$. Então N estará em $T^+(n)$ e $B = N \cdot U$. Por outro lado, se forem dados elementos N e U tais que N esteja em $T^+(n)$, U em $U(n)$ e $B = N \cdot U$, então $N^{-1}B$ estará em $U(n)$ e N^{-1} será a única matriz M caracterizada pelo teorema; além disso U será, necessariamente, $N^{-1}B$.

Exemplo 28. Sejam x_1 e x_2 números reais tais que $x_1^2 + x_2^2 = 1$ e $x_1 \neq 0$. Seja

$$B = \begin{bmatrix} x_1 & x_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Aplicando o processo de Gram-Schmidt às linhas de B, obteremos os vetores

$$\begin{aligned}\alpha_1 &= (x_1, x_2, 0)\\ \alpha_2 &= (0, 1, 0) - x_2(x_1, x_2, 0)\\ &= x_1(-x_2, x_1, 0)\\ \alpha_3 &= (0, 0, 1).\end{aligned}$$

Seja U a matriz cujas linhas são $\alpha_1, (\alpha_2/x_1), \alpha_3$. Então, U é unitária e

$$U = \begin{bmatrix} x_1 & x_2 & 0 \\ -x_2 & x_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\dfrac{x^2}{x_1} & \dfrac{1}{x_1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Multiplicando agora pela inversa de

$$M = \begin{bmatrix} 1 & 0 & 0 \\ -\dfrac{x_2}{x_1} & \dfrac{1}{x_1} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

vemos que

$$\begin{bmatrix} x_1 & x_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ x_2 & x_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & 0 \\ -x_2 & x_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Consideremos agora abreviadamente uma mudança de coordenadas em um espaço com produto interno. Suponhamos que V seja um espaço de dimensão finita com produto interno e que $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ e $\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ sejam duas bases *ortonormais* ordenadas de V. Existe uma única $n \times n$ matriz P (necessariamente inversível) tal que

$$[\alpha]_{\mathscr{B}} = P^{-1}[\alpha]_{\mathscr{B}}$$

para todo α em V. Se U é o bem determinado operador linear sobre V definido por $U\alpha_j = \alpha'_j$, então P é a matriz de U em relação à base ordenada $\mathscr{B}$.

$$\alpha'_k = \sum_{j=1}^{n} P_{jk}\alpha_j.$$

como $\mathscr{B}$ e $\mathscr{B}'$ são bases ortonormais, U é um operador unitário e P é uma matriz unitária. Se T é um operador linear arbitrário sobre V, então

$$[T]_{\mathscr{B}'} = P^{-1}[T]_{\mathscr{B}}P = P^*[T]_{\mathscr{B}}P.$$

Definição. *Sejam* A *e* B n $\times$ n *matrizes complexas. Dizemos que* B *é* **unitariamente equivalente** *a* A *se existe uma* n $\times$ n *matriz unitária* P *tal que* B $=$ P^{-1*}AP. *Dizemos que* B *é* **ortogonalmente equivalente** *a* A *se existe uma* n $\times$ n *matriz ortogonal* P *tal que* B $=$ P^{-1}AP.

Com essa definição, o que observamos acima pode ser enunciado como segue. Se $\mathscr{B}$ e $\mathscr{B}'$ são duas bases ortonormais ordenadas de V, para todo operador linear T sobre V, a matriz $[T]_{\mathscr{B}'}$ é unitariamente equivalente à matriz $[T]_{\mathscr{B}}$. No caso de V ser um espaço real com produto interno, estas matrizes são ortogonalmente equivalentes, através de uma matriz ortogonal real.

Exercícios

1. Determinar uma matriz unitária que não seja ortogonal e determinar uma matriz ortogonal que não seja unitária.

2. Seja V o espaço das $n \times n$ matrizes complexas com o produto interno $(A|B) = \operatorname{tr}(AB^*)$. Para cada M em V, seja T_M o operador linear definido por $T_M(A) = MA$. Mostrar que T_M é unitário se, e somente se, M é uma matriz unitária.

3. Seja V o conjunto dos números complexos, considerado como um espaço vetorial *real.*

   (a) Mostrar que $(\alpha|\beta) = Re(\alpha\bar{\beta})$ define um produto interno sobre V.

   (b) Exibir um isomorfismo (de espaço com produto interno) de V em R^2 com o produto interno canônico.

   (c) Para cada γ em V, seja M_γ o operador linear sobre V definido por $M_\gamma(\alpha) = \gamma\alpha$. Mostrar que $(M_\gamma)^* = M_{\bar{\gamma}}$.

   (d) Para quais números complexos γ se tem $M\gamma$ auto-adjunto?

   (e) Para quais γ, M_γ é unitário?

   (f) Para quais γ, M_γ é positivo? (ver definição na pág. 363).

   (g) Qual é o $\det(M_\gamma)$?

   (h) Determinar a matriz de M_γ em relação à base $\{1, i\}$.

   (i) Se T é um operador linear sobre V, encontrar condições necessárias e suficientes para que T seja um M_γ.

   (j) Encontrar um operador unitário sobre V que não seja um M_γ.

4. Seja V o espaço R^2, com o produto interno canônico. Se U é um operador unitário sobre V, mostrar que a matriz de U em relação à base ordenada canônica é

$$\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{bmatrix}$$

para algum θ, $0 \leq \theta < 2\pi$. Seja U_θ o operador linear correspondente à primeira matriz, isto é, U_θ é uma rotação de um ângulo θ. Agora é possível convencer se de que todo operador unitário sobre V é uma rotação ou uma reflexão em relação ao eixo ε_1 seguida de uma rotação.

(a) O que é $U_\theta U_\phi$?

(b) Mostrar que $U^*_\theta = U_{-\theta}$.

(c) Seja ϕ um número real fixo e $\mathscr{B} = \{\alpha_1, \alpha_2\}$ a base ortonormal obtida girando $\{\varepsilon_1, \varepsilon_2\}$ de um ângulo ϕ, isto é, $\alpha_j = U_\phi \varepsilon_j$. Se θ é um outro número real, qual é a matriz de U_θ em relação à base ordenada $\mathscr{B}$?

5. Seja V o espaço R^3, com o produto interno canônico. Seja W o plano gerado por $\alpha = (1, 1, 1)$ e $\beta = (1, 1, -2)$. Seja U o operador linear definido geometricamente como segue: U é uma rotação de um ângulo θ, em torno de uma reta que passa pela origem e é ortogonal a W. Existem na verdade duas tais rotações: tome se uma delas. Determinar a matriz de U em relação à base ordenada canônica. (Eis um modo possível de se proceder: Determinar α_1 e α_2 que formem uma base ortonormal de W. Seja α_3 um vetor de norma 1, ortogonal a W. Determinar a matriz de U em relação à base $\{\alpha_1, \alpha_2, \alpha_3\}$. Efetuar uma mudança de base.)

6. Seja V um espaço de dimensão finita com produto interno e seja W um subespaço de V. Então $V = W \oplus W^\perp$, isto é, cada α em V pode ser expresso de um único modo sob a forma $\alpha = \beta + \gamma$, com β em W e γ em $W^\perp$. Definamos um operador linear U por $U\alpha = \beta - \gamma$.

(a) Demonstrar que U é auto adjunto e unitário.

(b) Se V é R^3 com o produto interno canônico e W é o subespaço gerado por $(1, 0, 1)$, encontrar a matriz de U em relação à base ordenada canônica.

7. Seja V um espaço *complexo* com produto interno e T um operador linear *auto-adjunto* sobre V. Mostrar que

(a) $\|\alpha + iT\alpha\| = \|\alpha - iT\alpha\|$ para todo α em V.

(b) $\alpha + iT\alpha = \beta + iT\beta$ se, e somente se, $\alpha = \beta$.

(c) $I + iT$ é não-singular.

(d) $I - iT$ é não-singular.

(e) Suponhamos que V seja de dimensão finita; demonstrar que

$$U = (I - iT)(I + iT)^{-1}$$

é um operador unitário; U é denominado a **transformada de Cayley** de T. Num certo sentido, $U = f(T)$, onde $f(x) = (1 - ix)/(1 + ix)$.

8. Se θ é um número real, demonstrar que as matrizes seguintes são unitariamente equivalentes.

$$\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}, \qquad \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}.$$

9. Seja V um espaço de dimensão finita com produto interno e T um operador linear positivo sobre V. Seja p_T o produto interno sobre V definido por $p_T(\alpha, \beta) = (T\alpha|\beta)$. Seja U um operador linear sobre V e U^* o seu adjunto em relação a $(\,|\,)$. Demonstrar que U é unitário em relação ao produto interno p_T se, e somente se, $T = U^*TU$.

10. Seja V um espaço de dimensão finita com produto interno. Para cada par α, β em V, seja $T_{\alpha,\beta}$ o operador linear sobre V definido por $T_{\alpha,\beta}(\gamma) = (\gamma|\beta)\alpha$. Mostrar que

(a) $T^*_{\alpha,\beta} = T_{\beta,\alpha}$;

(b) traço $(T_{\alpha,\beta}) = (\alpha|\beta)$;

(c) $T_{\alpha,\beta}\, T_{\gamma,\delta} = T_{\alpha,(\beta|\gamma)\delta}$;

(d) Em que condições $T_{\alpha,\beta}$ é auto-adjunto?

11. Seja V um espaço n dimensional com produto interno sobre o corpo F e seja $L(V, V)$ o espaço dos operadores lineares sobre V. Mostrar que existe um único produto interno sobre $L(V, V)$ com a propriedade que $\|T_{\alpha,\beta}\|^2 = \|\alpha\|^2\,\|\beta\|^2$ para todos α, β em V. ($T_{\alpha,\beta}$ é o operador definido no Exercício 10.) Encontrar um isomorfismo entre $L(V, V)$ com este produto interno e o espaço das $n \times n$ matrizes sobre F, com o produto interno $(A|B) = \operatorname{tr}(AB^*)$.

12. Seja V um espaço de dimensão finita com produto interno. No Exercício 6, mostramos como construir certos operadores lineares sobre V que são auto-adjuntos e unitários. Demonstrar então que não existem outros, isto é, que todo operador auto adjunto unitário provém de algum subespaço W como descrevemos no Exercício 6.

13. Sejam V e W espaços de mesma dimensão finita com produto interno. Seja U um isomorfismo de V em W. Mostrar que

(a) a aplicação $T \to UTU^{-1}$ é um isomorfismo do espaço vetorial $L(V, V)$ no espaço vetorial $L(W, W)$;

(b) traço (UTU^{-1}) = traço (T) para todo T em $L(V, V)$;

(c) $UT_{\alpha,\beta}U^{-1} = T_{U\alpha,\, U\beta}$ ($T_{\alpha,\beta}$ definido no Exercício 10);

(d) $(UTU^{-1})^* = UT^*U^{-1}$;

(e) se equiparmos $L(V, V)$ com o produto interno $(T_1|T_2) = \text{traço}(T_1T_2^*)$, e analogamente para $L(W, W)$, então $T \to UTU^{-1}$ é um isomorfismo de espaço com produto interno.

14. Se V é um espaço com produto interno, um **movimento rígido** é uma função qualquer T de V em V (não necessariamente linear) tal que $\|T\alpha - T\beta\| = \|\alpha - \beta\|$

para todos α, β em V. Um exemplo de um movimento rígido é um operador linear unitário. Outro exemplo é uma translação por um vetor fixo γ.

$$T_\gamma(\alpha) = \alpha + \gamma$$

(a) Seja V o espaço R^2 com o produto interno canônico. Suponhamos que T seja um movimento rígido de V e que $T(0) = 0$. Demonstrar que T é linear e é um operador unitário.

(b) Usar o resultado da parte (a) para demonstrar que todo movimento rígido de R^2 é composto de uma translação, seguida de um operador unitário.

(c) Mostrar agora que um movimento rígido de R^2 é uma translação seguida de uma rotação ou então uma translação seguida de uma reflexão seguida de uma rotação.

15. Um operador unitário sobre R^4 (com o produto interno canônico) é simplesmente um operador linear que conserva a forma quadrática

$$\|(x, y, z, t)\|^2 = x^2 + y^2 + z^2 + t^2$$

isto é, um operador linear U tal que $\|U\alpha\|^2 = \|\alpha\|^2$ para todo α em R^4. Numa certa parte da teoria da relatividade, é de interesse determinar os operadores lineares T que conservam a forma

$$\|(x, y, z, t)\|_L^2 = t^2 - x^2 - y^2 - z^2.$$

Contudo, $\| \;\|_L^2$ não vem de um produto interno, mas de algo chamado "métrica de Lorentz" (a qual não estudaremos). Por esta razão, um operador linear T sobre R^4 tal que $\|T\alpha\|_L^2$, para todo α em R^4, é denominado uma **transformação de Lorentz.**

(a) Mostrar que a função U definida por

$$U(x, y, z, t) = \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

é um isomorfismo de R^4 no espaço vetorial H das 2×2 matrizes complexas auto-adjuntas.

(b) Mostrar que $\|\alpha\|_L^2 = \det(U\alpha)$.

(c) Suponhamos que T seja um operador linear (real) sobre o espaço H das 2×2 matrizes auto-adjuntas. Mostrar que $L = U^{-1}TU$ é um operador linear sobre R^4.

(d) Seja M uma 2×2 matriz complexa arbitrária. Mostrar que $T_M(A) = M^*AM$ define um operador linear T_M sobre H. (É necessário verificar que T_M leva H em H.)

(e) Se M é uma 2×2 matriz tal que $|\det M| = 1$, mostrar que $L_M = U^{-1}T_MU$ é uma transformação de Lorentz sobre R^4.

(*f*) Encontrar uma transformação de Lorentz que não seja uma L_M.

8.5 Operadores Normais

O objetivo principal desta seção é a resolução do problema seguinte: Se T é um operador linear sobre um espaço V de dimensão finita com produto interno, sob que condições V possui uma base ortonormal formada por vetores característicos de T? Em outras palavras, quando é que existe uma base *ortonormal* $\mathcal{B}$ de V, tal que a matriz de T em relação à base $\mathcal{B}$ seja diagonal?

Vamos iniciar deduzindo algumas condições necessárias sobre T, que mostraremos subseqüentemente serem suficientes. Suponhamos que $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$ seja uma base ortonormal de V com a propriedade

(8-16) $$T\alpha_j = c_j\alpha_j, \quad j = 1, \dots, n.$$

Isto diz simplesmente que a matriz de T em relação à base ordenada $\mathcal{B}$ é a matriz diagonal com elementos diagonais $c_1, \dots, c_n$. O operador adjunto T^* é representado em relação a esta mesma base ordenada pela matriz transposta conjugada, isto é, a matriz diagonal com elementos diagonais $\bar{c}_1, \dots, \bar{c}_n$. Se V é um espaço real com produto interno, os escalares $c_1, \dots, c_n$ são (evidentemente) reais e então temos $T = T^*$. Em outras palavras, se V é um espaço *real* de dimensão finita com produto interno e T é um operador linear para o qual existe uma base ortonormal de vetores característicos, então T deve ser auto-adjunto. Se V é um espaço complexo com produto interno, os escalares $c_1, \dots, c_n$ não são necessariamente reais, isto é, T não é necessariamente auto-adjunto. Mas notemos que T deve satisfazer

(8-17) $$TT^* = T^*T$$

De fato, duas matrizes diagonais quaisquer comutam e como T e T^* são ambas representadas por matrizes diagonais em relação à base ordenada $\mathcal{B}$, temos (8-17). É um fato bastante notável que esta condição também seja suficiente para implicar a existência de uma base ortonormal formada por vetores característicos.

Definição. *Seja* V *um espaço de dimensão finita com produto interno e* T *um operador linear sobre* V. *Dizemos que* T *é* **normal** *se comuta com seu adjunto, isto é,* TT* = T*T.

Todo operador auto-adjunto é normal, como também o é todo operador unitário. Todo múltiplo escalar de um operador

normal é normal; contudo, somas e produtos de operadores normais não são em geral normais. Embora isso não seja de forma alguma necessário, iniciaremos nosso estudo de operadores normais considerando operadores auto-adjuntos.

Teorema 15. *Seja* V *um espaço com produto interno e* T *um operador linear auto-adjunto sobre* V. *Todo valor característico de* T *é real. Vetores característicos de* T *associados a valores característicos distintos são ortogonais.*

Demonstração. Suponhamos que c seja um valor característico de T, isto é, que $T\alpha = c\alpha$ para algum vetor não-nulo α. Então

$$\begin{aligned} c(\alpha|\alpha) &= (c\alpha|\alpha) \\ &= (T\alpha|\alpha) \\ &= (\alpha|T\alpha) \\ &= (\alpha|c\alpha) \\ &= \bar{c}(\alpha|\alpha). \end{aligned}$$

Como $(\alpha|\alpha) \neq 0$, devemos ter $c = \bar{c}$. Suponhamos também que $T\beta = d\beta$ com $\beta \neq 0$. Então

$$\begin{aligned} c(\alpha|\beta) &= (T\alpha|\beta) \\ &= (\alpha|T\beta) \\ &= (\beta|d\beta) \\ &= \bar{d}(\alpha|\beta) \\ &= d(\alpha|\beta). \end{aligned}$$

Se $c \neq d$, então $(\alpha|\beta) = 0$.

Deve-se salientar que o Teorema 15 nada diz a respeito da existência de valores característicos ou de vetores característicos.

Teorema 16. *Em um espaço de dimensão finita positiva com produto interno, todo operador auto-adjunto possui um vetor característico (não-nulo).*

Demonstração. Seja V um espaço de dimensão n com produto interno, sendo $n > 0$ e seja T um operador auto-adjunto sobre V. Tomemos uma base ortonormal $\mathscr{B}$ de V e seja $A = [T]_{\mathscr{B}}$. Como $T = T^*$ temos $A = A^*$. Seja agora W o espaço das $n \times 1$ matrizes sobre C, com produto interno $(X|Y) = Y^*X$. Então $U(X) = AX$ define um operador linear auto-adjunto U sobre W. O polinômio característico, $\det(xI - A)$, é um polinômio de grau n sobre o corpo dos números complexos; todo polinômio sobre C de grau positivo

possui uma raiz. Assim, existe um número complexo c tal que $\det(cI - A) = 0$. Isto significa que $A - cI$ é singular, ou que existe uma X não-nula tal que $AX = cX$. Como o operador U (multiplicação por A) é auto-adjunto, decorre do Teorema 15 que c é real. Se V é um espaço vetorial real, podemos tomar X com elementos reais. De fato, nesse caso A e $A - cI$ têm elementos reais e como $A - cI$ é singular, o sistema $(A - cI)X = 0$ possui uma solução real não-nula X. Decorre que existe um vetor não-nulo α em V tal que $T\alpha = c\alpha$.

Diversos comentários devem ser feitos a respeito da demonstração.

(1) A demonstração da existência de um X não-nulo tal que $AX = cX$ nada teve que ver com o fato de A ser hermitiana (auto-adjunta). Ela mostra que todo operador linear sobre um espaço vetorial complexo de dimensão finita possui um vetor característico. No caso de um espaço real com produto interno, a auto-adjunção de A é usada de modo fundamental para nos dizer que cada valor característico de A é real e que, portanto, podemos encontrar um X conveniente com valores reais.

(2) O argumento mostra que o polinômio característico de uma matriz auto-adjunta tem coeficientes reais, a despeito do fato de que a matriz possa não ter elementos reais.

(3) A hipótese de V ser de dimensão finita é necessária para o teorema; um operador auto-adjunto sobre um espaço de dimensão infinita com produto interno pode não ter nenhum valor característico.

Exemplo 29. Seja V o espaço vetorial das funções complexas (ou reais) contínuas, definidas sobre o intervalo unitário $0 \leq t \leq 1$, com o produto interno

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\,dt.$$

O operador "multiplicação por t", $(Tf)(t) = tf(t)$, é auto-adjunto. Suponhamos que $Tf = cf$. Então

$$(t - c)f(t) = 0, \quad 0 \leq t \leq 1$$

e, então, $f(t) = 0$ para $t \neq c$. Como f é contínua, $f = 0$. Logo T não possui valores (vetores) característicos.

Teorema 17. *Seja* V *um espaço de dimensão finita com produto interno e seja* T *um operador linear arbitrário sobre* V. *Suponhamos que* W *seja um subespaço de* V *que seja invariante sob* T. *Então o suplementar ortogonal de* W *é invariante sob* T*.

Demonstração. Recordamos que o fato de W ser invariante sob T não quer dizer que cada vetor em W permaneça fixo por meio de T; significa que se α está em W então $T\alpha$ está em W. Seja β em $W^{\perp}$. Precisamos mostrar que $T^*\beta$ está em $W^{\perp}$, isto é, que $(\alpha|T^*\beta) = 0$ para todo α em W. Se α está em W, então $T\alpha$ está em W, portanto, $(T\alpha|\beta) = 0$. Mas $(T\alpha|\beta) = (\alpha|T^*\beta)^*$.

Teorema 18. *Seja* V *um espaço de dimensão finita com produto interno e seja* T *um operador linear auto-adjunto sobre* V. *Então existe uma base ortonormal de* V, *cujos vetores são vetores característicos de* T.

Demonstração. Estamos supondo $\dim V > 0$. Pelo Teorema 16, T possui um vetor característico α. Seja $\alpha_1 = \alpha/\|\alpha\|$ de modo que α_1 também é um vetor característico de T e $\|\alpha_1\| = 1$. Se $\dim V = 1$, já terminamos. Vamos agora proceder por indução sobre a dimensão de V. Suponhamos que o teorema seja válido para espaços com produto interno de dimensão menor que $\dim V$. Seja W o subespaço unidimensional gerado pelo vetor α_1. A afirmação de que α_1 é um vetor característico de T significa simplesmente que W é invariante sob T. Pelo Teorema 17, o suplementar ortogonal $W^{\perp}$ é invariante sob $T^* = T$. Ora $W^{\perp}$, com o produto interno de V, é um espaço com produto interno de dimensão um a menos que a dimensão de V. Seja U o operador linear induzido sobre $W^{\perp}$ por T, isto é, a restrição de T a $W^{\perp}$. Então U é auto-adjunto e, pela hipótese de indução, $W^{\perp}$ possui uma base ortonormal $\{\alpha_2, \dots, \alpha_n\}$ formada por vetores característicos de U. Ora, cada um desses vetores também é um vetor característico de T e como $V = W \oplus W^{\perp}$, concluímos que $\{\alpha_1, \dots, \alpha_n\}$ é a desejada base de V.

Corolário. *Seja* A *uma* n × n *matriz hermitiana* (*auto-adjunta*). *Então existe uma matriz unitária* P *tal que* $P^{-1}AP$ *seja diagonal* (A *é unitariamente equivalente a uma matriz diagonal*). *Se* A *é uma matriz simétrica real, existe uma matriz ortogonal real* P *tal que* $P^{-1}AP$ *seja diagonal.*

Demonstração. Seja V o espaço $C^{n\times 1}$, com o produto interno canônico e seja T o operador linear sobre V que é representado

por A em relação à base ordenada canônica. Como $A = A^*$, temos $T = T^*$. Seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base ortonormal ordenada de V, tal que $T\alpha_j = c_j\alpha_j$, $j = 1, \ldots, n$. Se $D = [T]_{\mathscr{B}}$, então D é a matriz diagonal com elementos diagonais $c_1, \ldots, c_n$. Seja P a matriz com vetores-colunas $\alpha_1, \ldots, \alpha_n$. Então $D = P^{-1}AP$.

Caso todo elemento de A seja real, podemos tomar V como sendo R^n, com o produto interno canônico e repetir o argumento. Neste caso, P será uma matriz unitária com elementos reais, ou seja, uma matriz ortogonal real.

Caso todo elemento de A seja real, podemos tomar V como sendo R^n, com o produto interno canônico e repetir o argumento. Neste caso, P será uma matriz unitária com elementos reais, ou seja, uma matriz ortogonal real.

Combinando o Teorema 18 com nossos comentários no início desta seção, temos o seguinte: Se V é um espaço *real* de dimensão finita com produto interno e T é um operador linear sobre V, então V possui uma base ortonormal formada por vetores característicos de T se, e somente se, T é auto-adjunto. Equivalentemente, se A é uma $n \times n$ matriz com elementos *reais*, existe uma matriz ortogonal real P tal que P^tAP seja diagonal se, e somente se, $A = A^t$. Não existe nenhum resultado semelhante para matrizes simétricas complexas. Em outras palavras, para matrizes complexas, existe uma diferença significativa entre as condições $A = A^t$ e $A = A^*$.

Tendo resolvido o caso de operadores auto-adjuntos, voltamos ao estudo dos operadores normais em geral. Vamos demonstrar o análogo do Teorema 18 para operadores normais, no caso *complexo*. Existe uma razão para essa restrição. Um operador normal sobre um espaço real com produto interno pode não ter nenhum vetor característico não nulo. Isso é verdade, por exemplo, para todas as rotações de R^2, salvo duas.

Teorema 19. *Seja* V *um espaço de dimensão finita com produto interno e seja* T *um operador normal sobre* V. *Suponhamos que* α *seja um vetor em* V. *Então* α *é um vetor característico de* T, *com valor característico* c *se, e somente se,* α *é um vetor característico de* T* *com valor característico* $\bar{c}$.

Demonstração. Suponhamos que U seja um operador normal arbitrário sobre V. Então $\|U\alpha\| = \|U^*\alpha\|$. De fato, usando a condição $UU^* = U^*U$, vemos que

$$\|U\alpha\|^2 = (U\alpha|U\alpha) = (\alpha|U^*U\alpha) =$$
$$= (\alpha|UU^*\alpha) = (U^*\alpha|U^*\alpha) = \|U^*\alpha\|^2.$$

Se c é um escalar arbitrário, o operador $U = T - cI$ é normal. De fato, $(T - cI)^* = T^* - \bar{c}I$ e é fácil verificar que $UU^* = U^*U$. Logo,

$$\|(T - cI)\alpha\| = \|(T^* - \bar{c}I)\alpha\|$$

de modo que $(T - cI)\alpha = 0$ se, e somente se, $(T^* - \bar{c}I)\alpha = 0$.

Definição. *Uma* n × n *matriz complexa* A *é dita* **normal** *se* $AA^* = A^*A$.

Não é muito fácil compreender o real significado de normalidade de matrizes ou operadores; no entanto, na sua tentativa de sentir o conceito, o leitor pode achar útil saber que uma matriz triangular é normal se, e somente se, ela é diagonal.

Teorema 20. *Sejam* V *um espaço de dimensão finita com produto interno,* T *um operador linear sobre* V *e* $\mathscr{B}$ *uma base ortonormal de* V. *Suponhamos que a matriz* A *de* T *em relação à base* $\mathscr{B}$ *seja triangular superior. Então* T *é normal se, e somente se,* A *é uma matriz diagonal.*

Demonstração. Como $\mathscr{B}$ é uma base ortonormal, A^* é a matriz de T^* em relação a $\mathscr{B}$. Se A é diagonal, então $AA^* = A^*A$ e isso implica $TT^* = T^*T$. Reciprocamente, suponhamos que T seja normal e $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$. Como A é triangular superior, $T\alpha_1 = A_{11}\alpha_1$. Usando o Teorema 19, isto implica $T^*\alpha_1 = \bar{A}_{11}\alpha_1$. Por outro lado,

$$T^*\alpha_1 = \sum_j (A^*)_{j1}\alpha_j$$
$$= \sum_j \bar{A}_{1j}\alpha_j.$$

Portanto, $A_{1j} = 0$ para todo $j > 1$. Em particular, $A_{12} = 0$ e como A é triangular superior, segue-se que

$$T\alpha_2 = A_{22}\alpha_2.$$

Assim, $T^*\alpha_2 = \bar{A}_{22}\alpha_2$ e $A_{2j} = 0$ para todo $j \neq 2$. Continuando dessa maneira, vemos que A é diagonal.

Teorema 21. *Seja* V *um espaço complexo de dimensão finita com produto interno e seja* T *um operador linear arbitrário sobre* V. *Então existe uma base ortonormal de* V *em relação à qual a matriz de* T *é triangular superior.*

Demonstração. Seja n a dimensão de V. O teorema é verdadeiro quando $n = 1$ e procederemos por indução sobre n, supondo que o resultado seja verdadeiro para operadores lineares sobre espaços complexos com produto interno, de dimensão $n - 1$. Como V é um espaço complexo de dimensão finita com produto interno, existe um vetor unitário α em V e um escalar c tal que

$$T^*\alpha = c\alpha.$$

Seja W o suplementar ortogonal do subespaço gerado por α e seja S a restrição de T a W. Pelo Teorema 17, W é invariante sob T. Logo, S é um operador linear sobre W. Como a dimensão de W é $n - 1$, nossa hipótese de indução implica a existência de uma base ortonormal $\{\alpha_1, \ldots, \alpha_{n-1}\}$ de W, em relação à qual a matriz de S é triangular superior: coloquemos $\alpha_n = \alpha$. Então $\{\alpha_1, \ldots, \alpha_n\}$ é uma base ortonormal de V em relação à qual a matriz de T é triangular superior.

Este teorema implica o seguinte resultado para matrizes:

Corolário. *Para toda* n × n *matriz complexa* A *existe uma matriz unitária* U *tal que* $U^{-1}AU$ *é triangular superior.*

Combinando, agora, os Teoremas 21 e 20, obtemos imediatamente o seguinte análogo ao Teorema 18, para operadores normais.

Teorema 22. *Seja* V *um espaço complexo de dimensão finita com produto interno e seja* T *um operador normal sobre* V. *Então* V *possui uma base ortonormal, onde cada vetor é um vetor característico de* T.

Novamente existe uma interpretação em termos de matrizes.

Corolário. *Para toda matriz normal* A *existe uma matriz unitária* P *tal que* $P^{-1}AP$ *é uma matriz diagonal.*

Exercícios

1. Para cada uma das seguintes matrizes simétricas reais A, encontrar uma matriz ortogonal real P tal que P^tAP seja diagonal.

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{bmatrix}$$

2. Uma matriz simétrica complexa é auto-adjunta? É normal?

3. Para

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

existe uma matriz ortogonal real P tal que $P^tAP = D$ seja diagonal. Determinar esta matriz diagonal D.

4. Seja V o espaço C^2, com o produto interno canônico. Seja T o operador linear sobre V que é representado em relação à base ordenada canônica pela matriz

$$A = \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}.$$

Mostrar que T é normal e determinar uma base ortonormal de V, constituída de vetores característicos de T.

5. Dar um exemplo de uma 2×2 matriz A tal que A^2 seja normal, mas A não seja normal.

6. Seja T um operador normal sobre um espaço complexo de dimensão finita com produto interno. Demonstrar que T é auto-adjunto, ou unitário conforme todo valor característico de T seja real, ou de valor absoluto 1. (Usar o Teorema 22 para reduzir o problema a um semelhante relativo a matrizes diagonais.)

7. Demonstrar que T é normal se, e somente se, $T = T_1 + iT_2$, onde T_1 e T_2 são operadores auto-adjuntos que comutam.

8. Demonstrar que uma matriz simétrica real possui uma raiz cúbica simétrica real, isto é, se A é simétrica real, existe uma B simétrica real tal que $B^3 = A$.

9. Demonstrar que se um operador é normal e nilpotente então ele é o operador nulo.

10. Se T é um operador normal, demonstrar que vetores característicos de T associados a valores característicos distintos são ortogonais.

11. Seja T um operador normal sobre um espaço complexo de dimensão finita com produto interno. Demonstrar que existe um polinômio f, com coeficientes complexos, tal que $T^* = f(t)$. (Representar T por meio de uma matriz diagonal e verificar o que f deve ser.)

12. Se dois operadores normais comutam, demonstrar que o seu produto é normal.

Capítulo 9

OPERADORES SOBRE ESPAÇOS COM PRODUTO INTERNO

9.1 Introdução

Consideramos os tópicos tratados no Capítulo 8 como fundamentais, fatos que todos deveriam conhecer. O presente capítulo se destina ao estudante mais avançado ou ao leitor ansioso de ampliar seus conhecimentos a respeito de operadores sobre espaços com produto interno. Com exceção do teorema do Eixo Principal, que é em essência apenas uma outra formulação do Teorema 18 sobre a diagonalização ortogonal de operadores auto-adjuntos, e dos outros resultados sobre formas na Seção 9.2, o material aqui apresentado é mais sofisticado e, em geral, mais envolvido tecnicamente. Também exigiremos mais do leitor, como já o fizemos nas últimas partes dos Capítulos 5 e 7. Os argumentos e demonstrações são escritos em um estilo mais condensado e não há quase exemplos que suavizem o caminho; tomamos a precaução, no entanto, de suprir o leitor com amplos conjuntos de exercícios.

As primeiras três seções se destinam a resultados ligados a formas sobre espaços com produto interno e relações entre formas e operadores lineares. A seção seguinte trata da teoria espectral, isto é, das implicações dos Teoremas 18 e 22 do Capítulo 9 concernentes à diagonalização de operadores auto-adjuntos e normais. Na seção final, prosseguimos o estudo de operadores normais tratando, em particular, o caso real e assim fazendo, examinamos o que o teorema da decomposição primária do Capítulo 6 diz a respeito de operadores normais.

9.2 Formas sobre Espaços com Produto Interno

Se T é um operador linear sobre um espaço V de dimensão finita com produto interno, a função f definida sobre $V \times V$ por

$$f(\alpha, \beta) = (T\alpha|\beta)$$

pode ser considerada como uma espécie de substituta de T. Muitas questões a respeito de T são equivalentes a questões a respeito de f. De fato, é fácil ver que f determina T, pois se $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ é uma base ortonormal de V, os elementos da matriz de T em relação a $\mathcal{B}$ são dados por

$$A_{jk} = f(\alpha_k, \alpha_j).$$

É importante entender, sob um ponto de vista mais abstrato, por que f determina T. As propriedades cruciais de f são descritas na seguinte definição.

Definição. *Uma* **forma (sesquilinear)** *sobre um espaço vetorial* V, *real ou complexo, é uma função sobre* $V \times V$ *com valores no corpo de escalares, tal que*

(a) $$f(c\alpha + \beta, \gamma) = cf(\alpha, \gamma) + f(\beta, \gamma)$$

(b) $$f(\alpha, c\beta + \gamma) = \bar{c}f(\alpha, \beta) + f(\alpha, \gamma)$$

para todos α, β, γ *em* V *e todos os escalares* c.

Assim, uma forma sesquilinear é uma função sobre $V \times V$ tal que $f(\alpha, \beta)$ é uma função linear de α, para β fixo e uma função linear-conjugada de β, para α fixo. No caso real, $f(\alpha, \beta)$ é linear como uma função de cada argumento; em outras palavras, f é uma **forma bilinear**. No caso complexo, a forma sesquilinear não é bilinear a não ser quando $f = 0$. No restante do capítulo omitiremos o adjetivo "sesquilinear" a não ser que pareça importante incluí-lo.

Se f e g são formas sobre V e c é um escalar, é fácil verificar que $cf + g$ também é uma forma. Segue daí que qualquer combinação linear de formas sobre V será novamente uma forma. Assim, o conjunto de todas as formas sobre V é um subespaço do espaço vetorial de todas as funções sobre $V \times V$ com valores escalares.

Teorema 1. *Seja* V *um espaço de dimensão finita com produto interno e* f *uma forma sobre* V. *Então existe um único operador linear* T *sobre* V *tal que*

$$f(\alpha, \beta) = (T\alpha|\beta)$$

para todos α, β *em* V *e a aplicação* $f \to T$ *é um isomorfismo do espaço das formas em* $L(V, V)$.

Demonstração. Fixemos um vetor β em V. Então $\alpha \to f(\alpha, \beta)$ é uma função linear sobre V. Pelo Teorema 6, existe um único vetor β' em V tal que $f(\alpha, \beta) = (\alpha|\beta')$ para todo α. Definamos uma função U de V em V, colocando $U\beta = \beta'$. Então

$$\begin{aligned} f(\alpha|c\beta + \gamma) &= (\alpha|U(c\beta + \gamma)) \\ &= \bar{c}f(\alpha, \beta) + f(\alpha, \gamma) \\ &= \bar{c}(\alpha|U\beta) + (\alpha|U\gamma) \\ &= (\alpha|cU\beta + U\gamma) \end{aligned}$$

para todos α, β, γ em V e todos os escalares c. Portanto, U é um operador linear sobre V e $T = U^*$ é um operador tal que $f(\alpha, \beta) = (T\alpha|\beta)$ para todos α e β. Se também tivermos $f(\alpha, \beta) = (T'\alpha|\beta)$, então

$$(T\alpha - T'\alpha|\beta) = 0$$

para todos α e β; assim $T\alpha = T'\alpha$ para todo α. Logo, para cada forma f existe um único operador linear T_f tal que

$$f(\alpha, \beta) = (T_f\alpha|\beta)$$

para todos α, β em V. Se f e g são formas e c um escalar,

$$\begin{aligned} (cf + g)(\alpha, \beta) &= (T_{cf+g}\alpha|\beta) \\ &= cf(\alpha, \beta) + g(\alpha, \beta) \\ &= c(T_f\alpha|\beta) + (T_g\alpha|\beta) \\ &= ((cT_f + T_g)\alpha|\beta) \end{aligned}$$

para todos α e β em V. Portanto,

$$T_{cf+g} = cT_f + T_g$$

de modo que $f \to T_f$ é uma aplicação linear. Para cada T em $L(V, V)$, a equação

$$f(\alpha, \beta) = (T\alpha|\beta)$$

define uma forma tal que $T_f = T$, e $T_f = 0$ se, e somente se, $f = 0$. Logo, $f \to T_f$ é um isomorfismo.

Corolário. *A equação*

$$(f|g) = tr(T_f T_g^*)$$

define um produto interno sobre o espaço das formas com a propriedade de que

$$(f|g) = \sum_{j,k} f(\alpha_k, \alpha_j)\overline{g(\alpha_k, \alpha_j)}$$

para toda base ortonormal $\{\alpha_1, \ldots, \alpha_n\}$ *de* **V**.

Demonstração. Do exemplo 3 do Capítulo 8, decorre facilmente que $(T, U) \to \mathrm{tr}(TU^*)$ é um produto interno sobre $L(V, V)$. Como $f \to T_f$ é um isomorfismo, o Exemplo 6 do Capítulo 8 mostra que

$$(f|g) = \mathrm{tr}(T_f T_g^*)$$

é um produto interno. Suponhamos, agora, que A e B sejam matrizes de T_f e T_g em relação à base ortonormal $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$. Então

$$A_{jk} = (T_f\alpha_k|\alpha_j) = f(\alpha_k, \alpha_j)$$

e $B_{jk} = (T_g\alpha_k|\alpha_j) = g(a_k, a_j)$. Como AB^* é a matriz de $T_f T_g^*$ em relação à base $\mathscr{B}$, segue-se que

$$(f|g) = \mathrm{tr}(AB^*) = \sum_{j,k} A_{jk}\overline{B}_{jk}.$$

Definição. *Se* f *é uma forma e* $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ *uma base ordenada arbitrária de* **V**, *a matriz* A *com elementos*

$$A_{jk} = f(\alpha_k, \alpha_j)$$

é denominada a **matriz** de f **em relação à base ordenada** $\mathscr{B}$.

Quando $\mathscr{B}$ for uma base ortonormal, a matriz de f em relação a $\mathscr{B}$ será também a matriz da transformação linear T_f mas, em geral, esse não é o caso.

Se A é a matriz de f em relação à base ordenada $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ segue-se que

(9-1) $$f(\sum_s x_s\alpha_s, \sum_r y_r\alpha_r) = \sum_{r,s} \bar{y}_r A_{rs} x_s$$

para todos os escalares x_s e $y_r (1 \leq r, s \leq n)$. Em outras palavras, a matriz A possui a propriedade de que

$$f(\alpha, \beta) = Y^*AX$$

onde X e Y são as respectivas matrizes das coordenadas de α e β em relação à base ordenada $\mathscr{B}$.

A matriz de f em relação a uma outra base

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i, \qquad (1 \leq j \leq n)$$

é dada pela equação

(9-2) $$A' = P^*AP.$$

De fato,

$$\begin{aligned} A'_{jk} &= f(\alpha'_k, \alpha'_j) \\ &= f(\sum_s P_{sk}\alpha_s, \sum_r P_{rj}\alpha_r) \\ &= \sum_{r,s} \overline{P_{rj}} A_{rs} P_{sk} \\ &= (P^*AP)_{jk}. \end{aligned}$$

Como para matrizes unitárias $P^* = P^{-1}$, decorre de (9-2) que no estudo de formas, podem ser aplicados os resultados concernentes à equivalência unitária.

Teorema 2. *Seja* f *uma forma sobre um espaço complexo* V *de dimensão finita com produto interno. Então existe uma base ortonormal de* V *em relação à qual a matriz de* f *é triangular superior.*

Demonstração. Seja T o operador linear sobre V tal que $f(\alpha, \beta) = (T\alpha|\beta)$ para todos α e β. Pelo Teorema 21, existe uma base ortonormal $\{\alpha_1, \ldots, \alpha_n\}$ em relação à qual a matriz de T é triangular superior. Logo,

$$f(\alpha_k, \alpha_j) = (T\alpha_k|\alpha_j) = 0$$

quando $j > k$.

Definição. *Uma forma* f *sobre um espaço vetorial real ou complexo é dita* **hermitiana** *se*

$$f(\alpha, \beta) = \overline{f(\beta, \alpha)}$$

para todos α *e* β *em* V.

Se T é um operador linear sobre um espaço V de dimensão finita com produto interno e f é a forma

$$f(\alpha, \beta) = (T\alpha|\beta)$$

então $\overline{f(\beta, \alpha)} = (\alpha|T\beta) = (T^*\alpha|\beta)$; portanto, f é hermitiana se, e somente se, T é auto-adjunto.

Quando f for hermitiana, $f(\alpha, \alpha)$ será real para todo α e, sobre espaços complexos, essa propriedade caracteriza as formas hermitianas.

Teorema 3. *Seja* V *um espaço vetorial complexo e* f *uma forma sobre* V *tal que* $f(\alpha, \alpha)$ *seja real para todo* α. *Então* f *é hermitiana.*

Demonstração. Sejam α e β vetores em V. Precisamos mostrar que $f(\alpha, \beta) = \overline{f(\beta, \alpha)}$. Ora,

$$f(\alpha + \beta, \alpha + \beta) = f(\alpha, \beta) + f(\alpha, \beta) + f(\beta, \alpha) + f(\beta, \beta).$$

Como $f(\alpha + \beta, \alpha + \beta)$, $f(\alpha, \alpha)$ e $f(\beta, \beta)$ são reais, o número $f(\alpha, \beta) + f(\beta, \alpha)$ é real. Examinando o mesmo argumento com $\alpha + i\beta$ no lugar de $\alpha + \beta$, vemos que $-if(\alpha, \beta) + if(\beta, \alpha)$ é real. Tendo concluído que dois números são reais, podemos igualá-los aos seus complexos conjugados e obter

$$f(\alpha, \beta) + f(\beta, \alpha) = \overline{f(\alpha, \beta)} + \overline{f(\beta, \alpha)}$$
$$-if(\alpha, \beta) + if(\beta, \alpha) = i\overline{f(\alpha, \beta)} - i\overline{f(\beta, \alpha)}$$

Se multiplicarmos a segunda equação por i e adicionarmos o resultado à primeira equação, obteremos

$$2f(\alpha, \beta) = 2\overline{f(\beta, \alpha)}.$$

Corolário. *Seja* T *um operador linear sobre um espaço complexo* V *de dimensão finita com produto interno. Então* T *é auto-adjunto se, e somente se,* $(T\alpha|\alpha)$ *é real para todo* α *em* V.

Teorema 4 (Teorema do Eixo Principal). *Para toda forma hermitiana* $\mathfrak{f}$ *sobre um espaço* V *de dimensão finita com produto interno, existe uma base ortonormal de* V *em relação à qual* $\mathfrak{f}$ *é representada por uma matriz diagonal com elementos reais.*

Demonstração. Seja T um operador linear tal que $f(\alpha, \beta) = (T\alpha|\beta)$ para todos α e β em V. Então, como $f(\alpha, \beta) = \overline{f(\beta, \alpha)}$ e $\overline{(T\beta|\alpha)} = (\alpha|T\beta)$, segue-se que

$$(T\alpha|\beta) = \overline{f(\beta, \alpha)} = (\alpha|T\beta)$$

para todos α e β; logo, $T = T^*$. Pelo Teorema 18 do Capítulo 8, existe uma base ortonormal de V constituída de vetores característicos de T. Suponhamos que $\{\alpha_1, \ldots, \alpha_n\}$ seja uma base ortonormal e que

$$T\alpha_j = c_j\alpha_j$$

para $1 \leq j \leq n$. Então,

$$f(\alpha_k, \alpha_j) = (T\alpha_k|\alpha_j) = \delta_{kj}c_k$$

e pelo Teorema 15 do Capítulo 8, cada c_k é real.

Corolário. *Sob as condições acima*

$$\mathfrak{f}(\sum_j x_j\alpha_j, \sum_k y_k\alpha_k) = \sum_j c_j x_j \bar{y}_j.$$

Exercícios

1. Quais das seguintes funções f, definidas sobre vetores $\alpha = (x_1, x_2)$ e $\beta = (y_1, y_2)$ de C^2, são formas (sesquilineares) sobre C^2?
   (a) $f(\alpha, \beta) = 1$.
   (b) $f(\alpha, \beta) = (x_1 - \bar{y}_1)^2 + x_2\bar{y}_2$.
   (c) $f(\alpha, \beta) = (x_1 + \bar{y}_1)^2 - (x_1 - \bar{y}_1)^2$.
   (d) $f(\alpha, \beta) = x_1\bar{y}_2 - \bar{x}_2 y_1$.

2. Seja f a forma sobre R^2 definida por

$$f((x_1, y_1), (x_2, y_2)) = x_1y_1 + x_2y_2.$$

Determinar a matriz de f em relação a cada uma das seguintes bases:

$$\{(1, 0), (0, 1)\}, \{(1, -1), (1, 1)\}, \{(1, 2), (3, 4)\}.$$

3. Seja

$$A = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}$$

e seja g a forma (sobre o espaço das 2×1 matrizes complexas) definida por $g(X, Y) = Y^*AX$. Verificar se g é um produto interno.

4. Seja V um espaço vetorial complexo e seja f uma forma (sesquilinear) sobre V que seja simétrica: $f(\alpha, \beta) = f(\beta, \alpha)$. Como é f?

5. Seja f a forma sobre R^2 dada por

$$f((x_1, x_2), (y_1, y_2)) = x_1y_1 + 4x_2y_2 + 2x_1y_2 + 2x_2y_1.$$

Encontrar uma base ordenada em relação à qual f seja representada por uma matriz diagonal.

6. Dizemos que a forma f é **não-degenerada** (à esquerda) se 0 for o único vetor α tal que $f(\alpha, \beta) = 0$ para todo β. Seja f uma forma sobre um espaço V com produto interno. Demonstrar que f é não-degenerada se, e somente se, o operador linear associado T_f (Teorema 1) é não singular.

7. Seja f uma forma sobre um espaço vetorial V de dimensão finita. Ver a definição de forma não-degenerada à esquerda, dada no Exercício 6. Definir forma não-degenerada à direita e demonstrar que a forma f é não-degenerada à esquerda se, e somente se, f é não-degenerada à direita.

8. Seja f uma forma não-degenerada (Exercícios 6 e 7) sobre um espaço V de dimensão finita. Seja L um funcional linear sobre V. Mostrar que existe um, e somente um, vetor β em V tal que $L(\alpha) = f(\alpha, \beta)$ para todo α.

9. Seja f uma forma não degenerada sobre um espaço V de dimensão finita. Mostrar que cada operador linear S possui um "adjunto relativo a f", isto é, um operador S' tal que $f(S\alpha, \beta) = f(\alpha, S'\beta)$ para todos α, β.

9.3 Formas Positivas

Nessa seção vamos discutir formas (sesquilineares) não-negativas e a sua relação com um dado produto interno sobre o espaço vetorial subjacente.

Definições. *Uma forma* f *sobre um espaço vetorial real ou complexo* V *é* **não-negativa** *se* f *é hermitiana e* $f(\alpha, \alpha) \geq 0$ *para todo* α *em* V. *A forma* f *é* **positiva** *se* f *é hermitiana e* $f(\alpha, \alpha) > 0$ *para todo* $\alpha \neq 0$.

Uma forma positiva sobre V é simplesmente um produto interno sobre V. Uma forma não-negativa satisfaz todas as proprie-

dades de um produto interno salvo que alguns vetores não nulos podem ser "ortogonais" a si mesmos.

Seja f uma forma sobre o espaço V de dimensão finita. Seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base ordenada de V e seja A a matriz de f em relação à base $\mathscr{B}$, isto é, $A_{jk} = f(\alpha_k, \alpha_j)$. Se $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$, então

$$\begin{aligned} f(\alpha, \alpha) &= f\left(\sum_j x_j\alpha_j, \sum_k x_k\alpha_k\right) \\ &= \sum_j \sum_k x_j\bar{x}_k f(\alpha_j, \alpha_k) \\ &= \sum_j \sum_k A_{kj} x_j \bar{x}_k. \end{aligned}$$

Assim, vemos que f é não-negativa se, e somente se,

$$A = A^*$$

e

$$\sum_j \sum_k A_{kj} x_j \bar{x}_k \geq 0 \tag{9-3}$$

para todos os escalares $x_1, \ldots, x_n$. A fim de que f seja positiva, a desigualdade (9-3) terá que ser estrita para todo $(x_1, \ldots, x_n) \neq 0$. As condições que deduzimos afirmam que f é uma forma positiva sobre V se, e somente se, a função

$$g(X, Y) = Y^*AX$$

é uma forma positiva sobre o espaço das $n \times 1$ matrizes coluna sobre o corpo de escalares.

Teorema 5. *Seja* F *o corpo dos números reais ou o corpo dos números complexos. Seja* A *uma* n × n *matriz sobre* F. *A função* g *definida por*

$$g(X, Y) = Y^*AX \tag{9-4}$$

é uma forma positiva sobre o espaço $F^{n \times 1}$ *se, e somente se, existir uma* n × n *matriz inversível* P, *com elementos em* F, *tal que* A = P*P.

Demonstração. Para uma $n \times n$ matriz A arbitrária, a função g em (9-4) é uma forma sobre o espaço das matrizes coluna. Estamos

tentando demonstrar que g é positiva se, e somente se, $A = P^*P$. Inicialmente, suponhamos $A = P^*P$. Então g é hermitiana e

$$\begin{aligned} g(X, X) &= X^*P^*PX \\ &= (PX)^*PX \\ &\geq 0. \end{aligned}$$

Se P for inversível e $X \neq 0$, então $(PX)^*PX > 0$.

Suponhamos, agora, que g seja uma forma positiva sobre o espaço das matrizes coluna. Então g é um produto interno e, portanto, existem matrizes coluna $Q_1, \ldots, Q_n$ tais que

$$\begin{aligned} \delta_{jk} &= g(Q_j, Q_k) \\ &= Q_k^* A Q_j. \end{aligned}$$

Mas isso apenas diz que se Q for a matriz com colunas $Q_1, \ldots, Q_n$ então $Q^*AQ = I$. Como $\{Q_1, \ldots, Q_n\}$ é uma base, Q é inversível. Façamos $P = Q^{-1}$ e teremos $A = P^*P$.

Na prática não é fácil verificar se uma dada matriz A satisfaz o critério de positividade estabelecido acima. Uma conseqüência do último teorema é que se g é positiva então $\det A > 0$, porque $\det A = \det(P^*P) = \det P^* \det P = |\det P|^2$. O fato de que $\det A > 0$ não é de modo algum suficiente para garantir que g seja positiva; no entanto existem n determinantes associados com A que possuem a seguinte propriedade: Se $A = A^*$ e se cada um desses determinantes for positivo, então g será uma forma positiva.

Definição. *Seja* A *uma* n × n *matriz sobre o corpo* F. *Os* **menores principais** *de* A *são os escalares* $\Delta_k(A)$ *definidos por*

$$\Delta_k(A) = det \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{k1} & \cdots & A_{kk} \end{bmatrix}, \quad 1 \leq k \leq n.$$

Lema. *Seja* A *uma* n × n *matriz inversível com elementos de um corpo* F. *As seguintes afirmações são equivalentes*:

(a) *Existe uma matriz triangular superior* P *com* $P_{kk} = 1$ $(1 \leq k \leq n)$ *tal que a matriz* $B = AP$ *é triangular inferior.*

(b) *Os menores principais de* A *são todos diferentes de* 0.

Demonstração. Seja P uma $n \times n$ matriz e façamos $B = AP$. Então

$$B_{jk} = \sum_r A_{jr} P_{rk}.$$

Se P é triangular superior e $P_{kk} = 1$ para todo k, então

$$\sum_{r=1}^{k-1} A_{jr} P_{rk} = B_{jk} - A_{kk}, \qquad k > 1.$$

Ora, B é triangular inferior desde que $B_{jk} = 0$ para $j < k$. Assim, B será triangular inferior se, e somente se,

$$\text{(9-5)} \qquad \sum_{r=1}^{k-1} A_{jr} P_{rk} = -A_{kk}, \qquad \begin{matrix} 1 \leq j \leq k - 1 \\ 2 \leq k \leq n. \end{matrix}$$

Vemos, assim, que a afirmação (a) do lema é equivalente à afirmação de que existem escalares P_{rk}, $1 \leq r \leq k$, $1 \leq k \leq n$, que satisfaçam (9-5) e $P_{kk} = 1$, $1 \leq k \leq n$.

Em (9-5), para cada $k > 1$ temos um sistema de $k - 1$ equações lineares nas incógnitas $P_{1k}, P_{2k}, \ldots, P_{k-1,k}$. A matriz dos coeficientes do sistema é

$$\begin{bmatrix} A_{11} & \cdots & A_{1,k-1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{k-1} & \cdots & A_{k-1,k-1} \end{bmatrix}$$

e seu determinante é o menor principal $\Delta_{k-1}(A)$. Se cada $\Delta_{k-1}(A) \neq 0$, os sistemas (9-5) possuem soluções únicas. Mostramos, assim, que a afirmação (b) implica a afirmação (a) e que a matriz P é única.

Suponhamos, agora, que (a) seja válida. Então, como veremos,

$$\text{(9-6)} \qquad \begin{aligned} \Delta_k(A) &= \Delta_k(B) \\ &= B_{11} B_{22}, \ldots, B_{kk}, \qquad k = 1, \ldots, n. \end{aligned}$$

Para verificar (9-6), sejam $A_1, \ldots, A_n$ e $B_1, \ldots, B_n$, respectivamente, as colunas de A e de B. Então

$$(9\text{-}7)\qquad \begin{aligned} B_1 &= A_1 \\ B_r &= \sum_{j=1}^{r-1} P_{jr}A_j + A_r, \qquad r > 1. \end{aligned}$$

Fixemos k, $1 \leq k \leq n$. Usando (9-7), vemos que a r-ésima coluna da matriz

$$\begin{bmatrix} B_{11} & \cdots & B_{kk} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ B_{k1} & \cdots & B_{kk} \end{bmatrix}$$

é obtida, adicionando-se à r-ésima coluna de

$$\begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{k1} & \cdots & A_{kk} \end{bmatrix}$$

uma combinação linear de suas outras colunas. Tais operações não alteram determinantes. Isso demonstra (9-6) a menos da observação trivial de que $\Delta_k(B) = B_{11}, \ldots, B_{kk}$ porque B é triangular. Como A e P são inversíveis, B é inversível. Portanto,

$$\Delta(B) = B_{11}, \ldots, B_{nn} \neq 0$$

e, assim, $\Delta_k(A) \neq 0$, $k = 1, \ldots, n$.

Teorema 6. *Seja* f *uma forma sobre um espaço vetorial* V *de dimensão finita e seja* A *a matriz de* f *em relação a uma base ordenada* $\mathcal{B}$. *Então* f *é uma forma positiva se, e somente se,* $A = A^*$ *e os menores principais de* A *são todos positivos.*

Demonstração. Provemos, inicialmente, a metade interessante do teorema. Suponhamos que $A = A^*$ e $\Delta_k(A) > 0$, $1 \leq k \leq n$. Pelo lema, existe uma (única) matriz triangular superior P com $P_{kk} = 1$, tal que $B = AP$ seja triangular inferior. A matriz P^* é triangular inferior, de modo que $P^*B = P^*AP$ também é triangular inferior.

Como A é auto-adjunta, a matriz $D = P^*AP$ é auto-adjunta. Uma matriz triangular auto-adjunta é necessariamente uma matriz diagonal. Por um raciocínio análogo ao usado para obter (9-6),

$$\begin{aligned}\Delta_k(D) &= \Delta_k(P^*B)\\ &= \Delta_k(B)\\ &= \Delta_k(A).\end{aligned}$$

Como D é diagonal, seus menores principais são

$$\Delta_k(D) = D_{11}, \ldots, D_{kk}.$$

De $\Delta_k(D) > 0$, $1 \leq k \leq n$, obtemos $D_{kk} > 0$ para cada k.

Se A é a matriz da forma f em relação à base ordenada $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$, $D = P^*AP$ é a matriz de f na base $\{\alpha', \ldots, \alpha'_n\}$ definida por

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Examinemos (9-2). Como D é diagonal, com elementos positivos na sua diagonal, é óbvio que

$$X^*DX > 0, \qquad X \neq 0$$

do que decorre que f é uma forma positiva.

Comecemos, agora, com uma forma positiva f. Sabemos que $A = A^*$. Como poderemos mostrar que $\Delta_k(A) > 0$, $1 \leq k \leq n$? Seja V_k o subespaço gerado por $\alpha_1, \ldots, \alpha_k$ e seja f_k a restrição de f a $V_k \times V_k$. Evidentemente f_k é uma forma positiva sobre V_k e, em relação à base $\{\alpha_1, \ldots, \alpha_k\}$, f_k é representada pela matriz

$$\begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{k1} & \cdots & A_{kk} \end{bmatrix}.$$

Já observamos, como conseqüência do Teorema 5, que o fato de que uma forma é positiva implica que o determinante de qualquer matriz que a representa é positivo.

Alguns comentários devem ser feitos para completar nosso estudo da relação existente entre formas positivas e matrizes. O que caracteriza as matrizes que representam formas positivas? Se f

é uma forma sobre um espaço vetorial complexo e A é a matriz de f em relação a alguma base ordenada, então f será positiva se, e somente se, $A = A^*$ e

(9-8) $$X^*AX > 0,$$

para toda matriz complexa $X \neq 0$. Decorre do Teorema 3 que a condição $A = A^*$ é redundante, isto é, (9-8) implica $A = A^*$. Por outro lado, se estivermos trabalhando com um espaço vetorial real, a forma f será positiva se, e somente se, $A = A^t$ e

(9-9) $$X^tAX > 0,$$

para toda matriz real $X \neq 0$. Queremos acentuar que do fato de que uma matriz real A satisfaça (9-9), não decorre que $A = A^t$. Um fato verdadeiro é o seguinte: se $A = A^t$ e (9-9) for válido então também (9-8) será válido. Isso porque

$$\begin{aligned}(X + iY)^*A(X + iY) &= (X^t - iY^t)A(X + iY)\\ &= X^tAX + Y^tAY + i[X^tAY - Y^tAX]\end{aligned}$$

e, se $A = A^t$, então $Y^tAX = X^tAY$.

Se A é uma $n \times n$ matriz com elementos complexos e se A satisfaz (9-9), diremos que A é uma **matriz positiva.** Os comentários que acabamos de fazer podem ser resumidos assim: Tanto no caso real como no complexo, uma forma f é positiva se, e somente se, sua matriz em relação a alguma (de fato, toda) base ordenada é uma matriz positiva.

Suponhamos, agora, que V seja um espaço de dimensão finita com produto interno. Seja f uma forma não-negativa sobre V. Existe um único operador linear auto-adjunto T sobre V tal que

(9-10) $$f(\alpha, \beta) = (T\alpha|\beta).$$

e T possui a propriedade adicional de que $(T\alpha|\alpha) \geq 0$.

Definição. *Um operador linear* T *sobre um espaço* V *de dimensão finita com produto interno é* **não-negativo**, *se* $T = T^*$ *e* $(T\alpha|\alpha) \geq 0$ *para todo* α *em* V. *Um operador linear* **positivo** é *um operador tal que* $T = T^*$ *e* $(T\alpha|\alpha) > 0$ *para todo* $\alpha \neq 0$.

Se V é um espaço vetorial (real ou complexo) de dimensão finita e se $(\cdot|\cdot)$ é um produto interno sobre V, então existe uma classe de operadores lineares positivos sobre V, associada a esse produto. Por intermédio de (9-10) existe uma correspondência bijetora entre essa classe de operadores positivos e a coleção de todas as formas positivas sobre V. Usaremos os exercícios dessa seção para acentuar as relações entre operadores positivos, formas positivas e matrizes positivas. O seguinte resumo poderá ser útil:

Se A é uma $n \times n$ matriz sobre o corpo dos números complexos, as seguintes afirmações são equivalentes:

(1) A é positiva, isto é, $\sum_j \sum_k A_{kj}x_j\bar{x}_k > 0$ sempre que $x_1, \ldots, x_n$ são números complexos, não todos nulos.

(2) $(X|Y) = Y^*AX$ é um produto interno sobre o espaço das $n \times 1$ matrizes complexas.

(3) Em relação ao produto interno canônico $(X|Y) = Y^*X$ sobre $n \times 1$ matrizes, o operador linear $X \to AX$ é positivo.

(4) $A = P^*P$ para alguma $n \times n$ matriz inversível P sobre C.

(5) $A = A^*$, e os menores principais de A são positivos.

Se todo elemento de A é real, estas são equivalentes a:

(6) $A = A^t$ e $\sum_j \sum_k A_{kj}x_jx_k > 0$ sempre que $x_1, \ldots, x_n$ são números reais não todos nulos.

(7) $(X|Y) = Y^tAX$ é um produto interno sobre o espaço das $n \times 1$ matrizes reais.

(8) Em relação ao produto interno canônico $(X|Y) = Y^tX$ sobre $n \times 1$ matrizes reais, o operador linear $X \to AX$ é positivo.

(9) Existe uma $n \times n$ matriz inversível P, com elementos reais, tal que $A = P^tP$.

Exercícios

1. Seja V igual a C^2, com o produto interno canônico. Para que vetores α em V existe um operador linear positivo T tal que $\alpha = T\varepsilon_1$?

2. Suponhamos que V seja R^2, com o produto interno canônico. Se θ é um número real, seja T_θ o operador linear "rotação de ângulo θ"

$$T_\theta(x_1, x_2) = (x_1 \cos\theta - x_2 \operatorname{sen}\theta,\ x_1 \operatorname{sen}\theta + x_2 \cos\theta).$$

Para que valores de θ se tem T_θ um operador positivo?

3. Seja V o espaço das $n \times 1$ matrizes sobre C, com o produto interno $(X|Y) = Y^* GX$ (onde G é uma $n \times n$ matriz tal que isto seja um produto interno). Seja A uma $n \times n$ matriz e T o operador linear $T(X) = AX$. Determinar T^*. Se Y é um elemento fixo de V, encontrar o elemento Z de V que determina o funcional linear $X \to Y^* X$. Em outras palavras, encontrar Z tal que $Y^* X = (X|Z)$ para toda X em V.

4. Seja V um espaço de dimensão finita com produto interno. Se T e U são operadores lineares positivos sobre V, demonstrar que $(T + U)$ é positivo. Dar um exemplo que mostre que TU não é necessariamente positivo.

5. Seja

$$A = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}.$$

(a) Mostrar que A é positiva.

(b) Seja V o espaço das 2×1 matrizes reais, com o produto interno $(X|Y) = Y^t AX$. Determinar uma base ortonormal de V, aplicando o processo de Gram-Schmidt à base $\{X_1, X_2\}$ definida por

$$X_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

(c) Determinar uma 2×2 matriz real P tal que $A = P^t P$.

6. Quais das matrizes são positivas?

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1+i \\ 1-i & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 1 \\ 3 & -1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

7. Dar um exemplo de uma $n \times n$ matriz cujos menores principais sejam todos positivos, mas que não seja uma matriz positiva.

8. Verificar se $((x_1, x_2)|(y_1, y_2)) = x_1\bar{y}_1 + 2x_2\bar{y}_1 + 2x_1\bar{y}_2 + x_2\bar{y}_2$ define um produto interno sobre C^2.

9. Demonstrar que todo elemento da diagonal principal de uma matriz positiva é positivo.

10. Seja V um espaço de dimensão finita com produto interno. Se T e U são operadores lineares sobre V, coloquemos $T < U$ se $U - T$ é um operador positivo. Demonstrar o seguinte:

(a) É impossível que $T < U$ e $U < T$.

(b) Se $T < U$ e $U < S$, então $T < S$.

(c) Se $T < U$ e $0 < S$, não é necessário que $ST < SU$.

11. Seja V um espaço de dimensão finita com produto interno e E a projeção ortogonal de V sobre algum subespaço.

(a) Demonstrar que, para todo número positivo arbitrário c, o operador $cI + E$ é positivo.

(b) Exprimir em termos de E um operador linear auto-adjunto T tal que $T^2 = I + E$.

12. Seja n um inteiro positivo e A a $n \times n$ matriz

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n-1} \end{bmatrix}$$

Demonstrar que A é positiva.

13. Seja A uma $n \times n$ matriz auto-adjunta. Demonstrar que existe um número real c tal que a matriz $cI + A$ seja positiva.

14. Demonstrar que o produto de dois operadores lineares é positivo se, e somente se, eles comutam.

15. Seja T um operador normal sobre um espaço complexo de dimensão finita com produto interno. Demonstrar que T é positivo se todo valor característico de T é positivo.

16. Seja T um operador linear sobre o espaço V de dimensão finita com produto interno e suponhamos que T seja positivo e unitário. Demonstrar que $T = I$.

17. Demonstrar que toda matriz positiva é o quadrado de uma matriz positiva.

9.4 Mais sobre Formas

Essa seção contém dois resultados que darão informações mais detalhadas sobre formas (sesquilineares).

Teorema 7. *Seja* f *uma forma sobre um espaço vetorial real ou complexo* V *e seja* $\{\alpha_1, \ldots, \alpha_r\}$ *uma base do subespaço* W *de dimensão finita de* V. *Seja* M *a* r × r *matriz com elementos*

$$M_{jk} = f(\alpha_k, \alpha_j)$$

e W' *o conjunto de todos os vetores* β *em* V *tais que* $f(\alpha, \beta) = 0$ *para todo* α *em* W. *Então* W' *é um subespaço de* V *e* $W \cap W' = \{0\}$ *se, e somente se,* M *é inversível. Quando esse for o caso,* $V = W + W'$.

Demonstração. Se β e γ são vetores em W' e c um escalar, então para todo α em W

$$\begin{aligned} f(\alpha, c\beta + \gamma) &= \bar{c} f(\alpha, \beta) + f(\alpha, \gamma) \\ &= 0. \end{aligned}$$

Logo, W' é um subespaço de V.

Suponhamos, agora, que $\alpha = \sum_{k=1}^{r} x_k \alpha_k$ e que $\beta = \sum_{j=1}^{r} y_j \alpha_j$. Então

$$\begin{aligned} f(\alpha, \beta) &= \sum_{j,k} \bar{y}_j M_{jk} x_k \\ &= \sum_k \left(\sum_j \bar{y}_j M_{jk} \right) x_k. \end{aligned}$$

Segue daí que $W \cap W' \neq \{0\}$ se, e somente se, o sistema homogêneo

$$\sum_{j=1}^{r} \bar{y}_j M_{jk} = 0, \qquad 1 \leq k \leq r$$

possuir uma solução não-trivial $(y_1, \ldots, y_r)$. Logo, $W \cap W' = \{0\}$ se, e somente se, M^* for inversível. Mas a inversibilidade de M^* é equivalente à inversibilidade de M.

Suponhamos que M seja inversível e coloquemos

$$A = (M^*)^{-1} = (M^{-1})^*.$$

Definamos g_j sobre V pela equação

$$g_j(\beta) = \sum_{k=1}^{r} A_{jk} \overline{f(\alpha_k, \beta)}.$$

Então

$$g_j(c\beta + \gamma) = \sum_k A_{jk} \overline{f(\alpha_k, c\beta + \gamma)}$$

$$= c \sum_k A_{jk} f(\alpha_k, \beta) + \sum_k A_{jk} f(\alpha_k, \gamma)$$
$$= c g_j(\beta) + g_j(\gamma).$$

Logo, cada g_j é uma função linear sobre V. Podemos, então, definir um operador linear E sobre V colocando

$$E\beta = \sum_{j=1}^{r} g_j(\beta)\alpha_j.$$

Como

$$g_j(\alpha_n) = \sum_k A_{jk}\overline{f(\alpha_k, \alpha_n)}$$
$$= \sum_k A_{jk}(M^*)_{kn}$$
$$= \delta_{jn}$$

segue-se que $E(\alpha_n) = \alpha_n$ para $1 \leq n \leq r$. Isso implica que $E\alpha = \alpha$ para todo α em W. Portanto, E leva V sobre W e $E^2 = E$. Se β é um vetor arbitrário em V, então

$$f(\alpha_n, E\beta) = f\left(\alpha_n, \sum_j g_j(\beta)\alpha_j\right)$$
$$= \sum_j \overline{g_j(\beta)}\, f(\alpha_n, \alpha_j)$$
$$= \sum_j \left(\sum_k \bar{A}_{jk} f(\alpha_k, \beta)\right) f(\alpha_n, \alpha_j).$$

Como $A^* = M^{-1}$, segue-se que

$$f(\alpha_n, E\beta) = \sum_k \left(\sum_j (M^{-1})_{kj} M_{jn}\right) f(\alpha_k, \beta)$$
$$= \sum_k \delta_{kn} f(\alpha_k, \beta)$$
$$= f(\alpha_n, \beta).$$

Isso implica que $f(\alpha, E\beta) = f(\alpha, \beta)$ para todo α em W. Logo,

$$f(\alpha, \beta - E\beta) = 0$$

para todo α em W e β em V. Assim, $I - E$ leva V em W'. A equação

$$\beta = E\beta + (I - E)\beta$$

mostra que $V = W + W'$. Um último fato deve ser mencionado. Como $W \cap W' = \{0\}$, todo vetor em V é, de modo único, soma de um vetor de W e um vetor de W'. Se β está em W', segue-se que $E\beta = 0$. Logo, $I - E$ leva V sobre W'.

A projeção E, construída na demonstração, pode ser caracterizada da seguinte maneira: $E\beta = \alpha$ se, e somente se, α estiver em W e $\beta - \alpha$ pertencer a W'. Portanto, E é independente da base de W que foi usada na sua construção. Podemos, então, nos referir a E como a **projeção de V sobre W**, determinada pela decomposição em soma direta

$$V = W \oplus W'.$$

Notemos que E é uma projeção ortogonal se, e somente se, $W' = W^{\perp}$.

Teorema 8. *Seja* f *uma forma sobre um espaço vetorial real ou complexo* V *e seja* A *a matriz de* f *em relação à base ordenada* $\{\alpha_1, \ldots, \alpha_n\}$ *de* V. *Suponhamos que os menores principais de* A *sejam todos diferentes de zero. Então existe uma única matriz triangular superior* P *com* $P_{kk} = 1$ $(1 \leq k \leq n)$ *tal que*

$$P^*AP$$

seja triangular superior.

Demonstração. Como $\Delta_k(A^*) = \overline{\Delta_k(A)}$ $(1 \leq k \leq n)$, os menores principais de A^* são todos diferentes de zero. Logo, de acordo com o lema usado na demonstração do Teorema 6, existe uma matriz triangular superior P com $P_{kk} = 1$ tal que A^*P seja triangular inferior. Portanto, $P^*A = (A^*P)^*$ é triangular superior. Como o produto de duas matrizes triangulares superiores é novamente triangular superior, segue-se que P^*AP é triangular superior. Isso mostra a existência de P mas não a sua unicidade. Existe, porém, um outro argumento mais geométrico, que pode ser usado para demonstrar tanto a existência como a unicidade de P.

Sejam W_k o subespaço gerado por $\alpha_1, \ldots, \alpha_k$ e W'_k o conjunto de todos os β em V tais que $f(\alpha, \beta) = 0$ para todo α em W_k. Como $\Delta_k(A) \neq 0$, a $k \times k$ matriz M com elementos

$$M_{ij} = f(\alpha_j, \alpha_i) = A_{ij}$$

$(1 \leq i, j \leq k)$ é inversível. Segundo o Teorema 7

$$V = W_k \oplus W'_k.$$

Seja E_k a projeção de V sobre W_k determinada por essa decomposição e façamos $E_0 = 0$. Seja

$$\beta_k = \alpha_k - E_{k-1}\alpha_k, \qquad (1 \leq k \leq n).$$

Então $\beta_1 = \alpha_1$ e $E_{k-1}\alpha_k$ pertence a W_{k-1} para $k > 1$. Assim, quando $k > 1$, existem escalares P_{jk}, determinados de modo único, tais que

$$E_{k-1}\alpha_k = -\sum_{j=1}^{k-1} P_{jk}\alpha_j.$$

Fazendo $P_{kk} = 1$ e $P_{jk} = 0$ para $j > k$, teremos uma $n \times n$ matriz triangular superior P com $P_{kk} = 1$ e

$$\beta_k = \sum_{j=1}^{k} P_{jk}\alpha_j$$

para $k = 1, \ldots, n$. Suponhamos $1 \leq i < k$. Então β_i está em W_i e $W_i \subset W_{k-1}$. Como β_k pertence a W'_{k-1}, segue-se que $f(\beta_i, \beta_k) = 0$. Indiquemos por B a matriz de f em relação à base ordenada $\{\beta_1, \ldots, \beta_n\}$. Então

$$B_{ki} = f(\beta_i, \beta_k)$$

de modo que $B_{ki} = 0$ para $k > i$. Portanto, B é triangular superior. Por outro lado,

$$B = P^*AP.$$

Reciprocamente, suponhamos que P seja uma matriz triangular superior com $P_{kk} = 1$ tal que P^*AP seja triangular superior. Coloquemos

$$\beta_k = \sum_j P_{jk}\alpha_j, \qquad (1 \leq k \leq n).$$

Então, evidentemente, $\{\beta_1, \ldots, \beta_k\}$ é uma base de W_k. Suponhamos $k > 1$. Então $\{\beta_1, \ldots, \beta_{k-1}\}$ é uma base de W_{k-1} e como $f(\beta_i, \beta_k) = 0$

quando $i < k$, vemos que β_k é um vetor de W'_{k-1}. A equação que define β_k implica

$$\alpha_k = -\left(\sum_{j=1}^{k-1} P_{jk}\alpha_j\right) + \beta_k.$$

Ora, $\sum_{j=1}^{k-1} P_{jk}\alpha_j$ pertence a W_{k-1} e β_k está em W'_{k-1}. Portanto, $P_{1k}, \ldots, P_{k-1k}$ são os únicos escalares tais que

$$E_{k-1}\alpha_k = -\sum_{j=1}^{k-1} P_{jk}\alpha_j$$

de modo que P é a matriz já construída anteriormente.

9.5 Teoria Espectral

Nessa seção desenvolveremos as implicações dos Teoremas 18 e 22 do Capítulo 8 no que diz respeito à diagonalização de operadores auto-adjuntos e normais.

Teorema 9 (**Teorema Espectral**). *Seja* T *um operador normal sobre um espaço complexo* V *de dimensão finita com produto interno ou um operador auto-adjunto sobre um espaço real* V *de dimensão finita com produto interno. Sejam* $c_1, \ldots, c_k$ *valores característicos distintos de* T. *Sejam* W_j *o espaço característico associado a* c_j *e* E_j *a projeção ortogonal de* V *sobre* W_j. *Então* W_j *é ortogonal a* W_i *quando* $i \neq j$, V *é a soma direta de* $W_1, \ldots, W_k$, *e*

$$T = c_1E_1 + \ldots + c_kE_k. \qquad \text{(9-11)}$$

Demonstração. Seja α um vetor em W_j, β um vetor em W_i e suponhamos $i \neq j$. Então

$$c_j(\alpha|\beta) = (T\alpha|\beta) = (\alpha|T^*\beta) = (\alpha|\bar{c}_i\beta).$$

Logo, $(c_j - c_i)(\alpha|\beta) = 0$ e como $c_j - c_i \neq 0$, segue-se que $(\alpha|\beta) = 0$. Assim, W_j é ortogonal a W_i quando $i \neq j$. Do fato de que V possui uma base ortonormal formada de vetores característicos (ver Teoremas 18 e 22 do Capítulo 8), segue-se que $V = W_1 + \ldots + W_k$. Se α_j pertence a V_j $(1 \leq j \leq k)$ e $\alpha_1 + \ldots + \alpha_k = 0$, então

$$0 = (\alpha_i | \sum_j \alpha_j) = \sum_j (\alpha_i | \alpha_j)$$
$$= \|\alpha_i\|^2$$

para todo i, de modo que V é a soma direta de $W_1, \ldots, W_k$. Portanto, $E_1 + \ldots + E_k = I$ e

$$T = TE_1 + \ldots + TE_k$$
$$= c_1E_1 + \ldots + c_kE_k.$$

A decomposição (9-11) é denominada a **resolução espectral** de T. Essa terminologia surgiu parcialmente de aplicações físicas que fizeram com que o espectro de um operador linear sobre um espaço vetorial de dimensão finita fosse definido como o conjunto de valores característicos do operador. É importante observar que as projeções ortogonais $E_1, \ldots, E_k$ são canonicamente associadas a T; na verdade elas são polinômios em T.

Corolário. *Se* $e_j = \prod_{i \neq j} \left(\frac{x - c_i}{c_j - c_i} \right)$, *então* $E_j = e_j(T)$ *para* $1 \leq j \leq k$.

Demonstração. Como $E_iE_j = 0$ quando $i \neq j$, segue-se que

$$T^2 = c_1^2E_1 + \ldots + c_k^2E_k$$

e por um simples argumento de indução, temos que

$$T^n = c_1^nE_1 + \ldots + c_k^nE_k$$

para todo inteiro $n \geq 0$. Para um polinômio arbitrário

$$f = \sum_{n=0}^{r} a_nx^n$$

temos

$$f(T) = \sum_{n=0}^{r} a_nT^n$$
$$= \sum_{n=0}^{r} a_n \sum_{j=1}^{k} c_j^nE_j$$

$$= \sum_{j=1}^{k} \left(\sum_{n=0}^{r} a_n c_j^n \right) E_j$$

$$= \sum_{j=1}^{k} f(c_j) E_j.$$

Como $e_j(c_m) = \delta_{jm}$, segue-se que $e_j(T) = E_j$.

Como $E_1, \ldots, E_k$ estão canonicamente associados a T e

$$I = E_1 + \ldots + E_k$$

a família das projeções $\{E_1, \ldots, E_k\}$ é denominada a **resolução do operador idêntico definida por** T.

Um comentário ainda deve ser feito a respeito da demonstração do teorema espectral. Deduzimos o teorema usando os Teoremas 18 e 22 do Capítulo 8, sobre a diagonalização de operadores auto-adjuntos e normais. Existe uma outra demonstração, mais algébrica, na qual deve ser demonstrado inicialmente que o polinômio minimal de um operador normal é um produto de fatores primos distintos. Procede-se então, como na demonstração do teorema da decomposição primária (Teorema 12, Capítulo 6). Daremos uma tal demonstração na próxima seção.

Em várias aplicações, é necessário saber se é, ou não, possível calcular certas funções de operadores ou matrizes, por exemplo, raízes quadradas. Isso pode ser feito de uma maneira bastante simples para operadores normais diagonalizáveis.

Definição. *Seja* T *um operador normal diagonalizável sobre um espaço de dimensão finita com produto interno e seja*

$$\mathrm{T} = \sum_{j=1}^{k} c_j E_j$$

sua resolução espectral. Suponhamos que f *seja uma função com valores no corpo de escalares e cujo domínio contenha o espectro de* T. *Então o operador linear* f(T) *é definido pela equação*

$$\text{(9-12)} \qquad f(\mathrm{T}) = \sum_{j=1}^{k} f(c_j) E_j.$$

Teorema 10. *Seja* T *um operador normal diagonalizável com espectro* S *sobre um espaço* V *de dimensão finita com produto interno. Suponhamos que* f *seja uma função com valores no corpo de escalares e cujo domínio contenha* S. *Então* f(T) *é um operador normal diagonalizável com espectro* f(S). *Se* U *é uma aplicação unitária de* V *sobre* V′ *e* $T' = UTU^{-1}$, *então* S *é o espectro de* T′ *e*

$$f(T') = Uf(T)U^{-1}.$$

Demonstração. A normalidade de $f(T)$ decorre de (9-12) por meio de um cálculo simples e do fato de que

$$f(T)^* = \sum_j \overline{f(c_j)} E_j.$$

Além disso, é claro que para todo α em $E_j(V)$

$$f(T)\alpha = f(c_j)\alpha.$$

Assim, o conjunto $f(S)$ de todos os $f(c)$ com c em S está contido no espectro de $f(T)$. Reciprocamente, suponhamos $\alpha \neq 0$ e que

$$f(T)\alpha = b\alpha.$$

Então $\alpha = \sum_j E_j\alpha$ e

$$\begin{aligned} f(T)\alpha &= \sum_j f(T)E_j\alpha \\ &= \sum_j f(c_j)E_j\alpha \\ &= \sum_j bE_j\alpha. \end{aligned}$$

Logo,

$$\begin{aligned} \|\sum_j (f(c_j) - b)E_j\alpha\|^2 &= \sum_j |f(c_j) - b|^2 \|E_j\alpha\|^2 \\ &= 0. \end{aligned}$$

Portanto, $f(c_j) = b$ ou $E_j\alpha = 0$. Por hipótese, $\alpha \neq 0$, de modo que existe um índice i tal que $E_i\alpha \neq 0$. Segue-se que $f(c_i) = b$ e portanto, que $f(S)$ é o espectro de $f(T)$. Suponhamos, de fato, que

$$f(S) = \{b_1, \ldots, b_r\}$$

com $b_m \neq b_n$ quando $m \neq n$. Seja X_m o conjunto dos índices i tais que $1 \leq i \leq k$ e $f(c_i) = b_m$. Seja $P_m = \sum_i E_i$, onde a somatória se estende sobre todos os índices i de X_m. Então P_m é a projeção ortogonal de V sobre o subespaço dos vetores característicos pertencentes ao valor característico b_m de $f(T)$, e

$$f(T) = \sum_{m=1}^{r} b_m P_m$$

é a resolução espectral de $f(T)$.

Suponhamos, agora, que U seja uma transformação unitária de V sobre V' e que $T' = UTU^{-1}$. Então a equação

$$T\alpha = c\alpha$$

é válida se, e somente se,

$$T'U\alpha = cU\alpha.$$

Assim, S é o espectro de T' e U transforma todo subespaço característico de T sobre o subespaço correspondente de T'. De fato, usando (9-12), vemos que

$$T' = \sum_j c_j E'_j, \qquad E'_j = UE_jU^{-1}$$

é a resolução espectral de T'. Logo,

$$\begin{aligned} f(T') &= \sum_j f(c_j) E'_j \\ &= \sum_j f(c_j) UE_jU^{-1} \\ &= U(\sum_j f(c_j) E_j) U^{-1} \\ &= Uf(T) U^{-1}. \end{aligned}$$

Examinando o que acabamos de ver, é importante manter em mente que o espectro do operador normal T é o conjunto

$$S = \{c_1, \ldots, c_k\}$$

de valores característicos distintos. Quando T é representado por uma matriz diagonal em relação a uma base de vetores característicos, torna-se necessário repetir cada valor c_j tantas vezes quanto for a dimensão do espaço de vetores característicos correspondente. Essa é a razão para a mudança de notação no seguinte resultado.

Corolário. *Com as hipóteses do Teorema* 10, *suponhamos que* T *seja representado, em relação à base ordenada* $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$, *pela matriz diagonal* D *com elementos* $d_1, \ldots, d_n$. *Então, em relação à base* $\mathscr{B}$, f(T) *é representado pela matriz diagonal* f(D) *com elementos* $f(d_1), \ldots, f(d_n)$. *Se* $\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ *é uma outra base ordenada e* P *a matriz tal que*

$$\alpha'_j = \sum_i P_{ij}\alpha_i$$

então $P^{-1}f(D)P$ *é a matriz de* f(T) *em relação à base* $\mathscr{B}'$.

Demonstração. Para cada índice i, existe um único j tal que $1 \leq j \leq k$, α_1 pertença a $E_j(V)$ e $d_i = c_j$. Portanto, $f(T)\alpha_i = f(d_i)\alpha_i$ para todo i, e

$$\begin{aligned} f(T)\alpha'_j &= \sum_i P_{ij} f(T)\alpha_i \\ &= \sum_i d_i P_{ij}\alpha_i \\ &= \sum_i (DP)_{ij}\alpha_i \\ &= \sum_i (DP)_{ij} \sum_k P^{-1}_{ki}\alpha'_k \\ &= \sum_k (P^{-1}DP)_{kj}\alpha'_k. \end{aligned}$$

Decorre desse resultado, que podemos formar certas funções de uma matriz normal. De fato, suponhamos que A seja uma matriz normal. Então existe uma matriz inversível P, na verdade uma matriz unitária P, tal que PAP^{-1} seja uma matriz diagonal, digamos D, com elementos $d_1, \ldots, d_n$. Seja f uma função assumindo valores complexos, que possa ser aplicada a $d_1, \ldots, d_n$ e seja $f(D)$ uma matriz diagonal com elementos $f(d_1), \ldots, f(d_n)$. Então $P^{-1}f(D)P$ é independente de D e somente uma função de A, no seguinte sentido. Se Q é uma outra matriz inversível tal que QAQ^{-1} seja uma

matriz diagonal D', então f pode ser aplicada aos elementos diagonais de D' e

$$P^{-1}f(D)P = Q^{-1}f(D')Q.$$

Definição. *Sob as condições acima,* f(A) *é definida como sendo* $P^{-1}f(D)P$.

A matriz $f(A)$ também pode ser caracterizada de uma maneira diferente. Para fazer isso, enunciaremos sem demonstração alguns resultados sobre matrizes normais, que são obtidos formulando-se, para matrizes, teoremas análogos aos anteriores.

Teorema 11. *Seja* A *uma matriz normal e* $c_1, \ldots, c_k$ *as raízes complexas distintas de* $\det(xI - A)$. *Seja*

$$e_i = \prod_{j \neq i} \left(\frac{x - c_j}{c_i - c_j} \right)$$

e $E_i = e_i(A)$ $(1 \leq i \leq k)$. *Então* $E_iE_j = 0$ *quando* $i \neq j$, $E_i^2 = E_i$, $E_i^* = E_i$, *e*

$$I = E_1 + \ldots + E_k.$$

Se f *é uma função com valores complexos cujo domínio inclua* $c_1, \ldots, c_k$, *então*

$$f(A) = f(c_1)E_1 + \ldots + f(c_k)E_k;$$

em particular,

$$A = c_1E_1 + \ldots + c_kE_k.$$

Lembremos que um operador T sobre um espaço com produto interno é não-negativo se T for auto-adjunto e $(T\alpha|\alpha) \geq 0$ para todo α em V.

Teorema 12. *Seja* T *um operador normal diagonalizável sobre um espaço* V *de dimensão finita com produto interno. Então* T *é auto-adjunto, não-negativo ou unitário, conforme cada valor característico de* T *seja real, não-negativo ou de valor absoluto* 1.

Demonstração. Suponhamos que T tenha a resolução espectral $T = c_1E_1 + \ldots + c_kE_k$. Então $T^* = \bar{c}_1E_1 + \ldots + \bar{c}_kE_k$. Dizer que T é auto-adjunto é dizer que $T = T^*$, ou

$$(c_1 - \overline{c}_1)E_1 + \ldots + (c_k - \overline{c}_k)E_k = 0.$$

Usando o fato de que $E_iE_j = 0$ para $i \neq j$ e o fato de que nenhum E_j é o operador nulo, vemos que T é auto-adjunto se, e somente se, $c_j = \overline{c}_j$, $j = 1, \ldots, k$. Para distinguir os operadores normais que são não-negativos, olhemos para

$$\begin{aligned}(T\alpha|\alpha) &= \left(\sum_{j=1}^{k} c_jE_j\alpha \middle| \sum_{i=1}^{k} E_i\alpha\right)\\ &= \sum_i \sum_j c_j(E_j\alpha|E_i\alpha)\\ &= \sum_j c_j \|E_j\alpha\|^2.\end{aligned}$$

Usamos o fato de que $(E_j\alpha|E_i\alpha) = 0$ para $i \neq j$. Fica claro então que a condição $(T\alpha|\alpha) \geq 0$ é satisfeita se, e somente se, $c_j \geq 0$ para cada j. Para distinguir os operadores unitários, observemos que

$$\begin{aligned}TT^* &= c_1c_1E_1 + \ldots + c_kc_kE_k\\ &= |c_1|^2 E_1 + \ldots + |c_k|^2 E_k.\end{aligned}$$

Se $TT^* = I$ então $I = |c_1|^2 E_1 + \ldots + |c_k|^2 E_k$, e, operando com E_j

$$E_j = |c_j|^2 E_j.$$

Como $E_j \neq 0$, temos $|c_j|^2 = 1$ ou $|c_j| = 1$. Reciprocamente, se $|c_j|^2 = 1$ para cada j, é claro que $TT^* = I$.

É importante notar que este é um teorema sobre operadores normais. Se T é um operador linear genérico sobre V que possui valores característicos reais, não decorre que T é auto-adjunto. O teorema afirma que se T possui valores característicos reais e *se* T é normal, então T é auto-adjunto. Um teorema deste tipo serve para fortalecer a analogia entre a operação de conjugação (operação de se tomar adjuntos) e o processo de se formar o conjugado de um número complexo. Um número complexo z é real ou de valor absoluto 1 conforme $z = \overline{z}$ ou $\overline{z}z = 1$. Um operador T é auto-adjunto ou unitário conforme $T = T^*$ ou $T^*T = I$.

Vamos, agora, demonstrar dois teoremas, que são os análogos destas duas afirmações.

(1) Todo número não-negativo possui uma única raiz quadrada não-negativa.

(2) Todo número complexo pode ser expresso sob a forma ru onde r é não-negativo e $|u| = 1$. Esta é a decomposição polar $z = re^{i\theta}$ para números complexos.

Teorema 13. *Seja* V *um espaço complexo de dimensão finita com produto interno e seja* T *um operador não-negativo sobre* V. *Então* T *possui uma única raiz quadrada não-negativa, isto é, existe um, e somente um, operador não-negativo* N *sobre* V *tal que* $N^2 = T$.

Demonstração. Seja $T = c_1E_1 + \ldots + c_kE_k$ a resolução espectral de T. Pelo Teorema 12, cada $c_j \geq 0$. Se c é um número real não-negativo arbitrário, indiquemos por $\sqrt{c}$ a raiz quadrada não-nega-negativa de c. Então, de acordo com o Teorema 11 e (9-12), $N = \sqrt{T}$ é um operador normal diagonalizável bem definido sobre V. Ele é não-negativo pelo Teorema 12 e, por um cálculo óbvio, $N^2 = T$.

Seja, agora, P um operador não-negativo sobre V tal que $p^2 = T$. Vamos demonstrar que $P = N$. Seja

$$P = d_1F_1 + \ldots + d_rF_r$$

a resolução espectral de P. Então $d_j \geq 0$ para cada j, pois P é não-negativo. De $P^2 = T$ resulta

$$T = d_1^2F_1 + \ldots + d_r^2F_r.$$

Ora, $F_1, \ldots, F_r$ satisfazem as condições $I = F_1 + \ldots + F_r$, $F_iJ_j = 0$ para $i \neq j$ e nenhum F_j é 0. Os números $d_1^2, \ldots, d_r^2$ são distintos, porque números não-negativos distintos possuem quadrados distintos. Pela unicidade da resolução espectral de T, devemos ter $r = k$ e (talvez reordenando) $F_j = E_j$, $d_j^2 = c_j$. Assim, $P = N$.

Teorema 14. *Seja* V *um espaço de dimensão finita com produto interno e seja* T *um operador linear arbitrário sobre* V. *Então, existe um operador unitário* U *sobre* V *e um operador não-negativo* N *sobre* V *tal que* $T = UN$. *O operador não-negativo* N *é único. Se* T *é inversível, o operador* U *também é único.*

Demonstração. Suponhamos que $T = UN$, onde U é unitário e N não-negativo. Então $T^* = (UN)^* = N^*U^* = NU^*$. Assim, $T^*T = NU^*UN = N^2$. Isto mostra que N é determinado de um

único modo como a raiz quadrada não-negativa do operador não-negativo T^*T.

Então, para iniciarmos a demonstração da existência de U e N, usaremos o Teorema 13 para definirmos N como a única raiz quadrada não-negativa de T^*T. Se T é inversível, então N também o é porque

$$(N\alpha|N\alpha) = (N^2\alpha|\alpha) = (T^*T\alpha|\alpha) = (T\alpha|T\alpha).$$

Neste caso, definamos $U = TN^{-1}$ e demonstraremos que U é unitário. Ora, $U^* = (TN^{-1})^* = (N^{-1})^*T^* = (N^*)^{-1}T^* = N^{-1}T^*$. Assim,

$$\begin{aligned} UU^* &= TN^{-1}N^{-1}T^* \\ &= T(N^{-1})^2T^* \\ &= T(N^2)^{-1}T^* \\ &= T(T^*T)^{-1}T^* \\ &= TT^{-1}(T^*)^{-1}T^* \\ &= I \end{aligned}$$

e U é unitário.

Se T é não inversível, teremos de realizar um pouco mais de trabalho para definir U. Definamos primeiro U sobre a imagem de N. Seja α um vetor na imagem de N, digamos, $\alpha = N\beta$. Definamos $U\alpha = T\beta$, motivados pelo fato de que queremos $UN\beta = T\beta$. Precisamos verificar que U está bem definido sobre a imagem de N; em outras palavras, se $N\beta' = N\beta$, então $T\beta' = T\beta$. Verificamos acima que $\|N\gamma\|^2 = \|T\gamma\|^2$ para todo γ em V. Assim, com $\gamma = \beta - \beta'$, vemos que $N(\beta - \beta') = 0$ se, e somente se, $T(\beta - \beta') = 0$. Portanto, U está bem definida sobre a imagem de N e é evidentemente linear onde definida. Se W é a imagem de N, vamos agora definir U sobre $W^\perp$. Para fazer isto precisamos da seguinte observação: Como T e N possuem o mesmo núcleo, suas imagens têm a mesma dimensão. Assim, $W^\perp$ possui mesma dimensão que o suplementar ortogonal da imagem de T. Portanto, existe um isomorfismo (de espaço com produto interno) U_0 de $W^\perp$ em $T(V)^\perp$. Agora já definimos U sobre W e vamos definir U sobre $W^\perp$ como sendo U_0.

Repitamos a definição de U. Como $V = W \oplus W^\perp$, cada α em V pode ser expresso de um único modo sob a forma $\alpha = N\beta + \gamma$, onde $N\beta$ está na imagem W de N e γ está em $W^\perp$. Definamos

$$U\alpha = T\beta + U_0\gamma.$$

Este U é evidentemente linear e, como verificamos acima, está bem definido. Além disso

$$\begin{aligned}(U\alpha|U\alpha) &= (T\beta + U_0\gamma|T\beta + U_0\gamma)\\ &= (T\beta|T\beta) + (U_0\gamma|U_0\gamma)\\ &= (N\beta|N\beta) + (\gamma|\gamma)\\ &= (\alpha|\alpha)\end{aligned}$$

portanto, U é unitário. Também $UN\beta = T\beta$ para cada β.

Denominamos $T = UN$ **a decomposição polar** de T. Certamente não dizemos que é a decomposição polar, pois U não é único. Mesmo quando T é inversível, de modo que U é único, temos a dificuldade de que U e N podem não comutar. Na verdade, eles comutam se, e somente se, T é normal. Por exemplo, se $T = UN = NU$, com N não-negativo e U unitário, então

$$TT^* = (NU)(NU)^* = NUU^*N = N^2 = T^*T.$$

O operador arbitrário T também possui uma decomposição $T = N_1U_1$, com N_1 não-negativo e U_1 unitário. Neste caso, N_1 será a raiz quadrada não-negativa de TT^*. Este resultado pode ser obtido aplicando-se o teorema há pouco demonstrado ao operador T^* e depois tomando adjuntos.

Voltaremos, agora, nossa atenção ao que pode ser dito a respeito da diagonalização simultânea de famílias comutativas de operadores normais. Para esse propósito a seguinte terminologia é apropriada.

Definições. *Seja* $\mathscr{F}$ *uma família de operadores sobre um espaço* V *com produto interno. Uma função* r *sobre* $\mathscr{F}$ *com valores no corpo* F *de escalares será denominada uma raiz de* $\mathscr{F}$ *se existir um* α *não nulo em* V *tal que*

$$\mathrm{T}\alpha = \mathrm{r}(\mathrm{T})\alpha$$

para todo T *em* $\mathscr{F}$. *Para uma função arbitrária* r *de* $\mathscr{F}$ *em* F, *indicaremos por* V(r) *o conjunto de todos os* α *em* V *tais que* $\mathrm{r}\alpha = \mathrm{r}(\mathrm{T})\alpha$ *para todo* T *em* $\mathscr{F}$.

Então $V(r)$ é um subespaço de V e r é uma raiz de $\mathscr{F}$ se, e somente se, $V(r) \neq \{0\}$. Todo α não nulo em $V(r)$ é simultaneamente um vetor característico para todo T em $\mathscr{F}$.

Teorema 15. *Seja* $\mathscr{F}$ *uma família comutativa de operadores normais diagonalizáveis sobre um espaço* V *de dimensão finita com produto interno. Então* $\mathscr{F}$ *possui somente um número finito de raízes. Se* $r_1, \ldots, r_k$, *são as raízes distintas de* $\mathscr{F}$, *então*

(i) $V(r_i)$ *é ortogonal a* $V(r_j)$ *se* $i \neq j$ *e*

(ii) $V = V(r_1) \oplus \ldots \oplus V(r_k)$.

Demonstração. Suponhamos que r e s sejam raízes distintas de $\mathscr{F}$. Então existe um operador T em $\mathscr{F}$ tal que $r(T) \neq s(T)$. Como vetores característicos pertencentes a valores característicos distintos de T são necessariamente ortogonais, segue-se que $V(r)$ é ortogonal a $V(s)$. Como V é de dimensão finita, $\mathscr{F}$ possui no máximo um número finito de raízes. Sejam $r_1, \ldots, r_k$ as raízes de $\mathscr{F}$. Suponhamos que $\{T_1, \ldots, T_m\}$ seja um subconjunto linearmente independente maximal de $\mathscr{F}$ e seja

$$\{E_{i1}, E_{i2}, \ldots\}$$

a resolução do operador idêntico definida por T_i $(1 \leq i \leq m)$. Então as projeções E_{ij} formam uma família comutativa. De fato, cada E_{ij} é um polinômio em T_i e $T_1, \ldots, T_m$ comutam entre si. Como

$$I = (\sum_{j_1} E_{1j_1}) (\sum_{j_2} E_{2j_2}) \ldots (\sum_{j_m} E_{mj_m})$$

cada vetor α em V pode ser escrito na forma

$$\alpha = \sum_{j_1, \ldots, j_m} E_{1j_1} E_{2j_2} \ldots E_{mj_m}\alpha. \tag{9-13}$$

Suponhamos que $j_1, \ldots, j_m$ sejam índices para os quais $\beta = E_{1j_1} E_{2j_2}, \ldots$ $\ldots, E_{mj_m}\alpha \neq 0$. Seja

$$\beta i = (\prod_{n \neq i} E_{nj_n})\alpha.$$

Então $\beta = E_{ij_i}\beta_i$; logo, existe um escalar c_i tal que

$$T_i\beta = c_i\beta, \qquad 1 \leq i \leq m.$$

Para cada T em $\mathscr{F}$, existem escalares b_i, determinados de modo único, tais que

$$T = \sum_{i=1}^{m} b_i T_i.$$

Assim,

$$T\beta = \sum_i b_i T_i \beta$$
$$= (\sum_i b_i c_i)\beta.$$

A função $T \to \sum_i b_i c_i$ é evidentemente uma das raízes, digamos r_t, de $\mathscr{F}$ e β está em $V(r_t)$. Portanto, cada termo não nulo em (9-13) pertence a um dos espaços $V(r_1), \ldots, V(r_k)$. Decorre daí que V é a soma direta ortogonal de $V(r_1), \ldots, V(r_k)$.

Corolário. *Sob as hipóteses do teorema, seja* P_j *a projeção ortogonal de* V *sobre* $V(r_j)$, $(1 \leq j \leq k)$. *Então* $P_iP_j = 0$ *se* $i \neq j$,

$$I = P_1 + \ldots + P_k,$$

e todo T *em* $\mathscr{F}$ *pode ser escrito na forma*

$$T = \sum_j r_j(T)P_j. \tag{9-14}$$

Definições. *A família das projeções ortogonais* $\{P_1, \ldots, P_k\}$ *é denominada a* **resolução do operador idêntico determinada por** $\mathscr{F}$, *e* (9-14) *é a* **resolução espectral de** T **em termos dessa família.**

Apesar de que as projeções $P_1, \ldots, P_k$, no corolário acima, estejam canonicamente associadas com a família $\mathscr{F}$, em geral elas não estão em $\mathscr{F}$ e nem são combinações lineares de operadores em $\mathscr{F}$; no entanto, mostraremos que elas podem ser obtidas formando-se certos produtos de polinômios em elementos de $\mathscr{F}$.

No estudo de uma família arbitrária de operadores lineares sobre um espaço com produto interno, é geralmente proveitoso considerar a álgebra auto-adjunta gerada por essa família.

Definição. *Uma* **álgebra auto-adjunta de operadores** *sobre um espaço* V *com produto interno é uma subálgebra linear de* $L(V, V)$ *que contém o adjunto de cada um de seus elementos.*

Um exemplo de uma álgebra auto-adjunta é o próprio $L(V, V)$. Como a interseção de qualquer coleção de álgebras auto-adjuntas é novamente uma álgebra auto-adjunta, a seguinte terminologia tem sentido.

Definição. *Se $\mathcal{F}$ é uma família de operadores lineares sobre um espaço de dimensão finita com produto interno, a* **álgebra auto-adjunta gerada por** $\mathcal{F}$ *é a menor álgebra auto-adjunta que contém $\mathcal{F}$.*

Teorema 16. *Seja $\mathcal{F}$ uma família comutativa de operadores normais diagonalizáveis sobre um espaço V de dimensão finita com produto interno, e seja $\mathcal{A}$ a álgebra auto-adjunta gerada por $\mathcal{F}$ e o operador idêntico. Seja $\{P_1, \ldots, P_k\}$ a resolução do operador idêntico definida por $\mathcal{F}$. Então $\mathcal{A}$ é o conjunto de todos os operadores sobre V da forma*

$$T = \sum_{j=1}^{k} c_j P_j \tag{9-15}$$

onde $c_1, \ldots, c_k$ são escalares arbitrários.

Demonstração. Indiquemos por $\mathcal{L}$ o conjunto de todos os operadores sobre V da forma (9-15). Então $\mathcal{L}$ contém o operador idêntico e o adjunto

$$T^* = \sum_j \bar{c}_j P_j$$

de cada um de seus elementos. Se $T = \sum_j c_j P_j$ e $U = \sum_j d_j P_j$, então para todo escalar a

$$aT + U = \sum_j (ac + d_j) P_j$$

e

$$\begin{aligned} TU &= \sum_{i,j} c_i d_j P_i P_j \\ &= \sum_j c_j d_j P_j \\ &= UT. \end{aligned}$$

Assim, $\mathcal{L}$ é uma álgebra comutativa auto-adjunta contendo $\mathcal{F}$ e o operador idêntico. Portanto, $\mathcal{L}$ contém $\mathcal{A}$.

Sejam $r_1, \ldots, r_k$ todas as raízes de $\mathcal{F}$. Então para cada par de índices (i, n), com $i \neq n$, existe um operador T_{in} em $\mathcal{F}$, tal que $r_i(T_{in}) \neq r_n(T_{in})$. Sejam $a_{in} = r_i(T_{in}) - r_n(T_{in})$ e $b_{in} = r_n(T_{in})$. Então o operador linear

$$Q_i = \prod_{n \neq i} a_{in}^{-1}(T_{in} - b_{in} I)$$

é um elemento da álgebra $\mathscr{A}$. Mostraremos que $Q_i = P_i (1 \leq i \leq k)$. Para esse fim, suponhamos $j \neq i$ e que α seja um vetor arbitrário em $V(r_j)$. Então

$$\begin{aligned} T_{ij}\alpha &= r_j(T_{ij})\alpha \\ &= b_{ij}\alpha \end{aligned}$$

de modo que $(T_{ij} - b_{ij}I)\alpha = 0$. Como todos os fatores de Q_i comutam, segue-se que $Q_i\alpha = 0$. Portanto, Q_i e P_i atuam da mesma maneira sobre $V(r_j)$ sempre que $j \neq i$. Suponhamos, agora, que α seja um vetor em $V(r_i)$. Então $T_{in}\alpha = r_i(T_{in})\alpha$, e

$$\mathscr{A}_{in}^{-1}(T_{in} - b_{in}I)\alpha = \mathscr{A}_{in}^{-1}[r_i(T_{in}) - r_n(T_{in})]\alpha = \alpha.$$

Assim, $Q_i\alpha = \alpha$ e Q_i e P_i atuam da mesma maneira sobre $V(r_i)$; portanto, $Q_i = P_i$ para $i = 1, \ldots, k$. Decorre disso que $\mathscr{A} = \mathfrak{L}$.

O teorema mostra que a álgebra $\mathscr{A}$ é comutativa e que cada elemento de $\mathscr{A}$ é um operador normal diagonalizável. Mostraremos a seguir que $\mathscr{A}$ possui um único gerador.

Corolário. *Sob as hipóteses do teorema, existe um operador* T *em* $\mathscr{A}$ *tal que todo elemento de* $\mathscr{A}$ *é um polinômio em* T.

Demonstração. Seja $T = \sum_{j=1}^{k} t_j P_j$, onde $t_1, \ldots, t_k$ são escalares distintos. Então

$$T^n = \sum_{j=1}^{k} t_j^n P_j$$

para $n = 1, 2, \ldots$. Se

$$f = \sum_{n=1}^{s} a_n x^n$$

segue-se que

$$\begin{aligned} f(T) = \sum_{n=1}^{s} a_n T^n &= \sum_{n=1}^{s} \sum_{j=1}^{k} a_n t_j^n P_j \\ &= \sum_{j=1}^{k} \left(\sum_{n=1}^{s} a_n t_j^n \right) P_j \\ &= \sum_{j=1}^{k} f(t_j) P_j. \end{aligned}$$

Dado um operador arbitrário

$$U = \sum_{j=1}^{k} c_j P_j$$

em $\mathscr{A}$, existe um polinômio f tal que $f(t_j) = c_j\,(1 \leq j \leq k)$ e para qualquer f nessas condições, $U = f(T)$.

Exercícios

1. Dar uma definição razoável de uma $n \times n$ matriz não-negativa e depois demonstrar que uma tal matriz possui uma única raiz quadrada não-negativa.

2. Seja A uma $n \times n$ matriz com elementos complexos tal que $A^* = -A$, e seja $B = e^A$. Mostrar que

   (a) $\det B = e^{\operatorname{tr} A}$;

   (b) $B^* = e^{-A}$;

   (c) B é unitária.

3. Se U e T são operadores normais que comutam, demonstrar que $U + T$ e UT são normais.

4. Seja T um operador linear sobre o espaço complexo V de dimensão finita com produto interno. Demonstrar que as dez afirmações seguintes sobre T são equivalentes:

   (a) T é normal.

   (b) $\|T\alpha\| = \|T^*\alpha\|$ para todos α em V.

   (c) $T = T_1 + iT_2$, onde T_1 e T_2 são auto-adjuntos e $T_1T_2 = T_2T_1$.

   (d) Se α é um vetor e c um escalar tal que $T\alpha = c\alpha$, então $T^*\alpha = \bar{c}\alpha$.

   (e) Existe uma base ortonormal de V formada por vetores característicos de T.

   (f) Existe uma base ortonormal $\mathscr{B}$ tal que $[T]_{\mathscr{B}}$ seja diagonal.

   (g) Existe um polinômio g com coeficientes complexos tal que $T^* = g(T)$.

   (h) Todo subespaço que é invariante sob T também é invariante sob T^*.

   (i) $T = NU$, onde N é não-negativo, U é unitário e N comuta com U.

   (j) $T = c_1E_1 + \ldots + c_kE_k$, onde $I = E_1 + \ldots + E_k$, $E_iE_j = 0$ para $i \neq j$, e $E_j^2 = E_j = E_j^*$.

5. Usar o Exercício 3 para mostrar que uma família arbitrária de operadores normais (não necessariamente diagonalizáveis) sobre um espaço de dimensão finita com produto interno, gera uma álgebra auto-adjunta comutativa de operadores normais.

6. Seja V um espaço de dimensão finita com produto interno e U um operador unitário sobre V tal que $U\alpha = \alpha$ implique $\alpha = 0$. Seja

$$f(z) = i\,\frac{(1+z)}{(1-z)}, \qquad z \neq 1$$

Mostrar que

(a) $f(U) = i(I+U)(I-U)^{-1}$;

(b) $f(U)$ é auto-adjunto;

(c) para todo operador auto-adjunto T sobre V, o operador

$$U = (T - iI)(T + iI)^{-1}$$

é unitário e tal que $T = f(U)$.

7. Seja V o espaço das $n \times n$ matrizes complexas, munido do produto interno

$$(A|B) = \operatorname{tr}(AB^*).$$

Se B é um elemento de V, indiquemos por L_B, R_B e T_B os operadores lineares sobre V definidos por

(a) $L_B(A) = BA$.

(b) $R_B(A) = AB$.

(c) $T_B(A) = BA - AB$.

Considerar as três famílias de operadores que se obtém, deixando B variar sobre todas as matrizes diagonais. Mostrar que cada uma dessas famílias é uma álgebra auto-adjunta comutativa e determinar suas resoluções espectrais.

8. Se B é um elemento arbitrário do espaço com produto interno do Exercício 7, mostrar que L_B é unitariamente equivalente a R_{B_t}.

9. Seja V o espaço com produto interno do Exercício 7 e G o grupo das matrizes unitárias em V. Se B está em G, indiquemos por C_B o operador linear sobre V, definido por

$$C_B(A) = BAB^{-1}.$$

Mostrar que

(a) C_B é um operador unitário sobre V;

(b) $C_{B_1B_2} = C_{B_1}C_{B_2}$;

(c) não existe nenhuma transformação unitária U sobre V tal que

$$UL_BU^{-1} = C_B$$

para todo B em G.

10. Seja $\mathscr{F}$ uma família arbitrária de operadores lineares sobre um espaço V de dimensão finita com produto interno e seja α a álgebra auto-adjunta gerada por $\mathscr{F}$. Mostrar que

(a) toda raiz de $\mathscr{A}$ define uma raiz de $\mathscr{F}$;

(b) toda raiz r de $\mathscr{A}$ é uma função linear multiplicativa sobre A, isto é,

$$r(TU) = r(T)r(U)$$

$$r(cT + U) = cr(T) + r(U)$$

para todos T e U em $\mathscr{A}$ e todos os escalares c.

11. Seja $\mathscr{F}$ uma família comutativa de operadores normais diagonalizáveis sobre um espaço V de dimensão finita com produto interno, e seja $\mathscr{A}$ a álgebra auto adjunta gerada por $\mathscr{F}$ e o operador idêntico I. Mostrar que cada raiz de $\mathscr{A}$ é diferente de 0, e que para cada raiz r de $\mathscr{F}$ existe uma única raiz s de $\mathscr{A}$ tal que $s(T) = r(T)$ para todo T em $\mathscr{F}$.

12. Seja $\mathscr{F}$ uma família comutativa de operadores normais diagonalizáveis sobre um espaço V de dimensão finita com produto interno e $\mathscr{A}_0$ a álgebra auto-adjunta gerada por $\mathscr{F}$. Seja $\mathscr{A}$ a álgebra auto-adjunta gerada por $\mathscr{F}$ e o operador idêntico I. Mostrar que

(a) $\mathscr{A}$ é o conjunto de todos os operadores sobre V da forma $cI + I$ sendo c um escalar e T um operador em $\mathscr{A}_0$.

(b) Existe no máximo uma raiz r de $\mathscr{A}$ tal que $r(T) = 0$ para todo T em $\mathscr{A}_0$.

(c Se uma das raízes de $\mathscr{A}$ é a função nula sobre α_0, as projeções $P_1, \ldots, P_k$ na resolução do operador idêntico definida por $\mathscr{F}$ podem ser indexadas de tal maneira que $\mathscr{A}_0$ seja constituída de todos os operadores sobre V da forma

$$T = \sum_{j=2}^{k} c_j P_j$$

onde $c_2, \ldots, c_k$ são escalares arbitrários.

(d) $\mathscr{A} = \mathscr{A}_0$ se, e somente se, para cada raiz r de $\mathscr{A}$ existir um operador T em $\mathscr{A}_0$ tal que $r(T) \neq 0$.

9.6 Outras Propriedades dos Operadores Normais

Na seção 8.5 desenvolvemos as propriedades fundamentais dos operadores auto-adjuntos e normais, usando os métodos mais simples e diretos possíveis. Na seção 9.5 consideramos vários aspectos da teoria espectral. Provaremos, agora, alguns resultados de natureza mais técnica a respeito, principalmente, de operadores normais sobre espaços reais.

Começaremos demonstrando uma versão mais profunda do teorema da decomposição primária, do Capítulo 6, para operadores normais. Ele se aplica tanto ao caso real como ao caso complexo.

Teorema 17. *Seja* T *um operador normal sobre um espaço* V *de dimensão finita com produto interno. Seja* p *o polinômio minimal de* T *e sejam* $p_1, \ldots, p_k$ *seus fatores primos unitários distintos. Então cada* p_j *ocorre com multiplicidade* 1 *na decomposição de* p *e possui grau* 1 *ou* 2. *Suponhamos que* W_j *seja o núcleo de* $p_j(T)$. *Então*

(i) W_j *é ortogonal a* W_i *se* $i \neq j$;

(ii) $V = W_1 \oplus, \ldots, \oplus W_k$;

(iii) W_j *é invariante sob* T *e* p_j *é o polinômio minimal da restrição de* T *a* W_j;

(iv) *para cada* j, *existe um polinômio* e_j *com coeficientes no corpo de escalares, tal que* $e_j(T)$ *é a projeção ortogonal de* V *sobre* W_j.

Na demonstração, usaremos certos fatos fundamentais que enunciaremos como lemas.

Lema 1. *Seja* N *um operador normal sobre um espaço* W *com produto interno. Então o núcleo de* N *é o suplementar ortogonal da sua imagem.*

Demonstração. Suponhamos $(\alpha|N\beta) = 0$ para todo β em W. Então $(N^*\alpha|\beta) = 0$ para todo β; logo $N^*\alpha = 0$. Pelo Teorema 19 do Capítulo 8, isso implica $N\alpha = 0$. Reciprocamente, se $N\alpha = 0$ então $N^*\alpha = 0$ e

$$(N^*\alpha|\beta) = (\alpha|N\beta) = 0$$

para todo β em W.

Lema 2. *Se* N *é um operador normal e* α *um vetor tal que* $N^2\alpha = 0$, *então* $N\alpha = 0$.

Demonstração. Suponhamos que N seja normal e que $N^2\alpha = 0$. Então $N\alpha$ está na imagem de N e também está no núcleo de N. Pelo Lema 1, isso implica $N\alpha = 0$.

Lema 3. *Seja* T *um operador normal e* f *um polinômio arbitrário com coeficientes no corpo de escalares. Então* f(T) *também é normal.*

Demonstração. Suponhamos $f = a_0 + a_1x + \ldots + a_nx^n$. Então

$$f(T) = a_0I + a_1T + \ldots + a_nT^n$$

e

$$f(T)^* = \bar{a}_0I + \bar{a}_1T^* + \ldots + \bar{a}_n(T^*)^n.$$

Como $T^*T = TT^*$, segue-se que $f(T)$ comuta com $f(T)^*$.

Lema 4. *Sejam* T *um operador normal e* f *e* g *polinômios relativamente primos, com coeficientes no corpo de escalares. Suponhamos*

que α *e* β *sejam vetores tais que* $f(T)\alpha = 0$ *e* $g(T)\beta = 0$. *Então* $(\alpha|\beta) = 0$.

Demonstração. Existem polinômios a e b, com coeficientes no corpo de escalares, tais que $af + bg = 1$. Assim,

$$a(T)f(T) + b(T)g(T) = I$$

e $\alpha = g(T)b(T)\alpha$. Segue-se que

$$(\alpha|\beta) = (g(T)b(T)\alpha|\beta) = (b(T)\alpha|g(T)^*\beta).$$

Por hipótese $g(T)\beta = 0$. Pelo Lema 3, $g(T)$ é normal. Portanto, pelo Teorema 19 do Capítulo 8, $g(T)^*\beta = 0$; logo $(\alpha|\beta) = 0$.

Demonstração do Teorema 17. Lembremos que o polinômio minimal de T é o polinômio unitário de menor grau entre todos os polinômios f tais que $f(T) = 0$. A existência de tais polinômios decorre da hipótese de que V é de dimensão finita. Suponhamos que algum fator primo p_j de p esteja repetido. Então $p = p_j^2 g$ para algum polinômio g. Como $p(T) = 0$, segue-se que

$$(p_j(T))^2 g(T)\alpha = 0$$

para todo α em V. Pelo Lema 3, $p_j(T)$ é normal. O Lema 2 implica

$$p_j(T)g(T)\alpha = 0$$

para todo α em V. Mas isso contradiz a hipótese de que p possua o menor grau entre todos os f tais que $f(T) = 0$. Portanto, $p = p_1, \ldots, p_k$. Se V é um espaço complexo com produto interno, cada p_j é necessariamente da forma

$$p_j = x - c_j$$

com c_j real ou complexo. Por outro lado, se V é um espaço real com produto interno, então $p_j = x_j - c_j$ com c_j em R, ou

$$p_j = (x - c)(x - \bar{c})$$

sendo c um número complexo não real.

Façamos, agora, $f_j = p/p_j$. Então, como $f_1, \ldots, f_k$ são relativamente primos, existem polinômios g_j com coeficientes no corpo de escalares, tais que

$$1 = \sum_j f_j g_j. \tag{9-16}$$

Indicaremos, resumidamente, como tais g_j podem ser construídos. Se $p_j = x - c_j$, então $f_j(c_j) \neq 0$, e para g_j tomamos o polinômio constante $1/f_j(c_j)$. Quando todos os p_j são dessa forma, os $f_j g_j$ são os conhecidos polinômios de Lagrange associados a $c_1, \ldots, c_k$, e (9-16) é evidentemente válido. Suponhamos algum $p_j = (x - c)(x - \bar{c})$ sendo c um número complexo não real. Então V é um espaço real com produto interno e tomamos

$$g_j = \frac{x - \bar{c}}{s} + \frac{x - c}{\bar{s}}$$

onde $s = (c - \bar{c}) f_j(c)$. Então

$$g_j = \frac{(s + \bar{s})x - (cs + \bar{c}s)}{s\bar{s}}$$

de modo que g_j é um polinômio com coeficientes reais. Se o grau de p é n, então

$$1 - \sum_j f_j g_j$$

é um polinômio com coeficientes reais de grau no máximo igual a $n - 1$; além disso, ele se anula para cada uma das n raízes (complexas) de p e é, portanto, identicamente nulo.

Seja α, agora, um vetor arbitrário em V. Então, por (9-16)

$$\alpha = \sum_j f_j(T)\, g_j(T)\alpha$$

e como $p_j(T) f_j(T) = 0$, segue-se que $f_j(T) g_j(T)\alpha$ está em W_j para todo j. Pelo Lema 4, W_j é ortogonal a W_i, sempre que $i \neq j$. Portanto, V é a soma direta ortogonal de $W_1, \ldots, W_k$. Se β é um vetor arbitrário em W_j,

$$p_j(T) T\beta = T p_j(T)\beta = 0;$$

logo W_j é invariante sob T. Seja T_j a restrição de T a W_j. Então $p_j(T_j) = 0$ de modo que p_j é divisível pelo polinômio minimal de T_j. Como p_j é irredutível sobre o corpo de escalares, segue-se que p_j é o polinômio minimal de T_j.

Em seguida, seja $e_j = f_j g_j$ e $E_j = e_j(T)$. Então, para todo vetor α em V, $E_j\alpha$ está em W_j, e

$$\alpha = \sum_j E_j\alpha.$$

Assim, $\alpha - E_i\alpha = \sum_{j \neq i} E_j\alpha$; como W_j é ortogonal a W_i quando $j \neq i$, isso implica que $\alpha - E_i\alpha$ está em $W_i^{\perp}$. Decorre agora, do Teorema 4 do Capítulo 8, que E_i é a projeção ortogonal de V sobre W_i.

Definição. *Os subespaços* W_j $(1 \leq j \leq k)$ *são denominados as* **componentes primárias de** V **sob** T.

Corolário. *Seja* T *um operador normal sobre um espaço* V *de dimensão finita com produto interno e sejam* $W_1, \ldots, W_k$ *as componentes primárias de* V *sob* T. *Suponhamos que* W *seja um subespaço de* V, *invariante sob* T. *Então*

$$W = \sum_j W \cap W_j.$$

Demonstração. Obviamente, W contém $\sum_j W \cap W_j$. Por outro lado, sendo W invariante sob T, ele será invariante sob qualquer polinômio em T. Em particular, W é invariante sob a projeção ortogonal E_j de V sobre W_j. Se α está em W, segue-se que $E_j\alpha$ está em $W \cap W_j$ e, ao mesmo tempo, $\alpha = \sum_j E_j\alpha$. Portanto, W está contido em $\sum_j W \cap W_j$.

O Teorema 17 mostra que todo operador normal T, sobre um espaço de dimensão finita com produto interno, é canonicamente especificado por um número finito de operadores normais T_j, definidos sobre as componentes primárias W_j de V sob T e cada um dos polinômios minimais desses operadores é irredutível sobre o corpo de escalares. Para compreender totalmente os operadores normais é necessário estudar os operadores normais desse tipo especial.

Um operador normal cujo polinômio minimal possui grau 1 é obviamente um múltiplo escalar do operador idêntico. Por outro lado, quando o polinômio minimal for irredutível e de grau 2 a situação torna-se mais complicada.

Exemplo 1. Suponhamos $r > 0$ e que θ seja um número real não múltiplo inteiro de π. Seja T o operador linear sobre R^2 cuja matriz em relação à base canônica é

$$A = r\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

Então T é um múltiplo escalar de uma transformação ortogonal e, portanto, é normal. Seja p o polinômio característico de T. Então

$$\begin{aligned} p &= \det(xI - A) \\ &= (x - r\cos\theta)^2 + r^2\operatorname{sen}^2\theta \\ &= x - 2r\cos\theta x + r^2. \end{aligned}$$

Seja $a = r\cos\theta$, $b = r\operatorname{sen}\theta$ e $c = a + ib$. Então $b \neq 0$, $c = re^{i\theta}$

$$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

e $p = (x - c)(x - \bar{c})$. Portanto, p é irredutível sobre R. Como p é divisível pelo polinômio minimal de T, segue-se que p é o polinômio minimal.

Esse exemplo sugere a seguinte recíproca.

Teorema 18. *Seja* T *um operador normal sobre um espaço* V *de dimensão finita com produto interno e* p *o seu polinômio minimal. Suponhamos que*

$$\mathrm{p} = (\mathrm{x} - \mathrm{a})^2 + \mathrm{b}^2$$

onde a *e* b *são reais e* $\mathrm{b} \neq 0$. *Então existe um inteiro* $\mathrm{s} > 0$ *tal que* p^{s} *é o polinômio característico de* T *e existem subespaços* $\mathrm{V}_1, \ldots, \mathrm{V}_{\mathrm{s}}$ *de* V *tais que*

(i) V_{j} *é ortogonal a* V_{i} *se* $\mathrm{i} \neq \mathrm{j}$;

(ii) $\mathrm{V} = \mathrm{V}_1 \oplus \ldots \oplus \mathrm{V}_{\mathrm{s}}$;

(iii) *cada* V_j *possui uma base ortonormal* $\{\alpha_j, \beta_j\}$ *com a propriedade de que*

$$T\alpha_j = a\alpha_j + b\beta_j$$
$$T\beta_j = -b\alpha_j + a\beta_j.$$

Em outras palavras, se $r = \sqrt{a^2 + b^2}$ e θ é escolhido de modo que $a = r\cos\theta$ e $b = r \text{ sen } \theta$, então V é uma soma direta ortogonal de subespaços V_j, de dimensão dois, sobre os quais T atua como "r vezes uma rotação de ângulo θ".

A demonstração do Teorema 18 se baseará no seguinte resultado.

Lema. *Seja* V *um espaço real com produto interno e* S *um operador normal sobre* V *tal que* $S^2 + I = 0$. *Seja* α *um vetor arbitrário de* V *e* $\beta = S\alpha$. *Então*

$$S^*\alpha = -\beta$$
(9-17)
$$S^*\beta = \alpha$$

$(\alpha|\beta) = 0$, *e* $\|\alpha\| = \|\beta\|$.

Demonstração. Temos $S\alpha = \beta$ e $S\beta = S^2\alpha = -\alpha$. Portanto,

$$0 = \|S\alpha - \beta\|^2 + \|S\beta + \alpha\|^2 = \|S\alpha\|^2 - 2(S\alpha|\beta) + \|\beta\|^2 + \|S\beta\|^2 + 2(S\beta|\alpha) + \|\alpha\|^2.$$

Como S é normal, segue-se que

$$0 = \|S^*\alpha\|^2 - 2(S^*\beta|\alpha) + \|\beta\|^2 + \|S^*\beta\|^2 + 2(S^*\alpha|\beta) + \|\alpha\|^2 = \|S^*\alpha + \beta\|^2 + \|S^*\beta - \alpha\|^2.$$

Isso implica (9-17); logo

$$(\alpha|\beta) = (S^*\beta|\beta) = (\beta|S\beta) = (\beta|-\alpha) = -(\alpha|\beta)$$

e $(\alpha|\beta) = 0$. Analogamente

$$\|\alpha\|^2 = (S^*\beta|\alpha) = (\beta|S\alpha) = \|\beta\|^2.$$

Demonstração do Teorema 18. Seja $V_1, \ldots, V_s$ uma coleção maximal de subespaços de dimensão dois, satisfazendo (i) e (ii) e as condições adicionais

$$\begin{aligned} T^*\alpha_j &= a\alpha_j - b\beta_j, \\ T^*\beta_j &= b\alpha_j + a\beta_j \end{aligned} \qquad 1 \leq j \leq s. \tag{9-18}$$

Seja $W = V_1 + \ldots + V_s$. Então W é a soma direta ortogonal de $V_1, \ldots, V_s$. Vamos mostrar que $W = V$. Suponhamos que isso não aconteça. Então $W^\perp \neq \{0\}$. Além disso, como (iii) e (9-18) implicam que W é invariante sob T e T^*, segue-se que $W^\perp$ é invariante sob T^* e $T = T^{**}$. Seja $S = b^{-1}(T - aI)$. Então $S^* = b^{-1}(T^* - aI)$, $S^*S = SS^*$ e $W^\perp$ é invariante sob S e S^*. Como $(T - aI)^2 + b^2I = 0$, segue-se que $S^2 + I = 0$. Seja α um vetor arbitrário de norma 1 em $W^\perp$ e façamos $\beta = S\alpha$. Então β está em $W^\perp$ e $S\beta = -\alpha$. Como $T = aI + bS$, isso implica

$$\begin{aligned} T\alpha &= a\alpha + b\beta \\ T\beta &= -b\alpha + a\beta. \end{aligned}$$

Pelo lema, $S^*\alpha = -\beta$, $S^*\beta = \alpha$, $(\alpha|\beta) = 0$ e $\|\beta\| = 1$. Porque $T^* = aI + bS^*$, segue-se que

$$\begin{aligned} T^*\alpha &= a\alpha + b\beta \\ T^*\beta &= b\alpha + \alpha\beta. \end{aligned}$$

Mas isso contradiz o fato de que $V_1, \ldots, V_s$ é uma coleção maximal de subespaços satisfazendo (i), (ii) e (9-18). Portanto, $W = V$ e como

$$\det \begin{bmatrix} x - a & b \\ -b & x - a \end{bmatrix} = (x - a)^2 + b^2$$

decorre de (i), (ii) e (iii) que

$$\det (xI - T) = [(x - a)^2 + b^2]^s.$$

Corolário. *Sob as condições do teorema,* T *é inversível e*

$$T^* = (a^2 + b^2)T^{-1}.$$

Demonstração. Como

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} = \begin{bmatrix} a^2 + b^2 & 0 \\ 0 & a^2 + b^2 \end{bmatrix}$$

segue-se de (iii) e (9-18) que $TT^* = (a^2 + b^2)I$. Logo, T é inversível e $T^* = (a^2 + b^2)T^{-1}$.

Teorema 19. *Seja* T *um operador normal sobre um espaço* V *de dimensão finita com produto interno. Então qualquer operador linear que comute com* T, *também comuta com* T*. *Além do mais, todo subespaço invariante sob* T *é também invariante sob* T*.

Demonstração. Suponhamos que U seja um operador linear sobre V que comute com T. Seja E_j a projeção ortogonal de V sobre a componente primária $W_j (1 \leq j \leq k)$ de V sob T. Então E_j é um polinômio em T e, portanto, comuta com U. Assim,

$$E_j U E_j = U E_j^2 = U E_j.$$

Portanto, $U(W_j)$ é um subconjunto de W_j. Indiquemos por T_j e U_j as restrições de T e U a W_j. Suponhamos que I_j seja o operador idêntico sobre W_j. Então U_j comuta com T_j e se $T_j = c_j I_j$, é claro que U_j também comuta com $T^*_j = \bar{c}_j I_j$. Por outro lado, se T_j não é um múltiplo escalar de I_j, então T_j é inversível e existem números reais a_j e b_j tais que

$$T_j^* = (a_j^2 + b_j^2) T_j^{-1}.$$

Como $U_j T_j = T_j U_j$, segue-se que $T_j^{-1} U_j = U_j T_j^{-1}$. Portanto, U_j comuta com T_j^* em ambos os casos. Ora, T^* também comuta com E_j e, portanto, W_j é invariante sob T^*. Além disso, para todos α e β em W_j

$$(T_j \alpha | \beta) = (T\alpha | \beta) = (\alpha | T^* \beta) = (\alpha | T_j^* \beta).$$

Como $T^*(W_j)$ está contido em W_j, isso implica que T_j^* é a restrição de T^* a W_j. Assim,

$$U T^* \alpha_j = T^* U \alpha_j$$

para todo α_j em W_j. Como V é a soma de $W_1, \ldots, W_k$, segue-se que

$$UT^*\alpha = T^*U\alpha$$

para todo α em V e portanto, U comuta com T^*.

Suponhamos, agora, que W seja um subespaço de V, invariante sob T, e seja $Z_j = W \cap W_j$. Pelo corolário do Teorema 17, $W = \sum_j Z_j$. Assim, é suficiente mostrar que cada Z_j é invariante sob T_j^*. Isso é óbvio se $T_j = c_j I$. Se esse não for o caso, T_j é inversível e transforma Z_j em, e, portanto, sobre Z_j. Logo, $T_j^{-1}(Z_j) = Z_j$ e como

$$T_j^* = (a_j^2 + b_j^2)T_j^{-1}$$

segue-se que $T^*(Z_j)$ está contido em Z_j para todo j.

Suponhamos que T seja um operador normal sobre um espaço V de dimensão finita com produto interno. Seja W um subespaço invariante sob T. Então, o corolário acima mostra que W é invariante sob T^*. Disso decorre que $W^\perp$ é invariante sob $T^{**} = T$ (e, portanto, sob T^* também). Usando esse fato, pode-se demonstrar facilmente a seguinte versão mais forte do teorema da decomposição cíclica dada no Capítulo 7.

Teorema 20. *Seja* T *um operador linear sobre um espaço* V *de dimensão finita com produto interno* (*dim* $V \geq 1$). *Então existem* r *vetores não nulos* $\alpha_1, \ldots, \alpha_r$ *em* V, *com respectivos* T-*anuladores* $e_1, \ldots, e_r$, *tais que*

(i) $V = Z(\alpha_1; T) \oplus \ldots \oplus Z(\alpha_r; T)$;

(ii) *se* $1 \leq k \leq r - 1$, *então* e_{k+1} *divide* e_k;

(iii) $Z(\alpha_j; T)$ *é ortogonal a* $Z(\alpha_k; T)$ *quando* $j \neq k$. *Além do mais, o inteiro* r *e os anuladores* $e_1, \ldots, e_r$ *são determinados de modo único pelas condições* (i) *e* (ii) *e pelo fato de que nenhum* α_k *é nulo.*

Corolário. *Se* A *é uma matriz normal com elementos reais* (*complexos*), *então existe uma matriz real ortogonal* (*unitária*) P *tal que* $P^{-1}AP$ *esteja sob a forma racional canônica.*

Segue daí que duas matrizes normais A e B são unitariamente equivalentes se, e somente se, elas possuem a mesma forma racio-

nal; A e B são ortogonalmente equivalentes se seus elementos forem reais e elas possuírem a mesma forma racional.

Por outro lado, existe um critério mais simples para a equivalência unitária de matrizes normais e operadores normais.

Definições. *Sejam* V *e* V′ *espaços com produto interno sobre o mesmo corpo. Uma transformação linear*

$$U : V \to V'$$

é denominada uma **transformação unitária** *se ela transforma* V *sobre* V′ *e preserva produtos internos. Se* T *é um operador linear sobre* V *e* T′ *é um operador linear sobre* V′, *então* T *é* **unitariamente equivalente** *a* T′ *se existir uma transformação unitária* U *de* V *sobre* V′ *tal que*

$$UTU^{-1} = T'.$$

Lema. *Sejam* V *e* V′ *espaços de dimensão finita com produto interno sobre o mesmo corpo. Suponhamos que* T *seja um operador linear sobre* V *e que* T′ *seja um operador linear sobre* V′. *Então* T *é unitariamente equivalente a* T′ *se, e somente se, existir uma base ortonormal* $\mathscr{B}$ *de* V *e uma base ortonormal* $\mathscr{B}'$ *de* V′ *tal que*

$$[T]_{\mathscr{B}} = [T']_{\mathscr{B}'}.$$

Demonstração. Suponhamos que exista uma transformação unitária U de V sobre V' tal que $UTU^{-1} = T'$. Seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base (ordenada) ortogonal arbitrária de V. Seja $\alpha'_j = U\alpha_j (1 \leq j \leq n)$. Então $\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ é uma base ortonormal de V' e colocando

$$T\alpha_j = \sum_{k=1}^{n} A_{kj}\alpha_k$$

vemos que

$$\begin{aligned} T'\alpha'_j &= UT\alpha_j \\ &= \sum_k A_{kj}U\alpha_k \\ &= \sum_k A_{kj}\alpha'_k \end{aligned}$$

Logo, $[T]_{\mathscr{B}} = A = [T']_{\mathscr{B}}$.

Reciprocamente, suponhamos que exista uma base ortonormal $\mathscr{B}$ de V e uma base ortonormal $\mathscr{B}'$ de V' tal que

$$[T]_{\mathscr{B}} = [T']_{\mathscr{B}}$$

e seja $A = [T]_{\mathscr{B}}$. Suponhamos $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ e $\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$. Seja U a transformação linear de V em V' tal que $U\alpha_j = \alpha'_j \, (1 \leq j \leq n)$. Então U é uma transformação unitária de V sobre V' e

$$\begin{aligned} UTU^{-1}\alpha'_j &= UT\alpha_j \\ &= U \sum_k A_{kj}\alpha_k \\ &= \sum_k A_{kj}\alpha'_k. \end{aligned}$$

Portanto, $UTU^{-1}\alpha'_j = T'\alpha'_j (1 \leq j \leq n)$, e isso implica que $UTU^{-1} = T'$.

Decorre imediatamente desse lema, que operadores unitariamente equivalentes sobre espaços de dimensão finita possuem o mesmo polinômio característico. Para operadores normais vale a recíproca.

Teorema 21. *Sejam* V *e* V′ *espaços de dimensão finita com produto interno sobre o mesmo corpo. Suponhamos que* T *seja um operador normal sobre* V *e que* T′ *seja um operador normal sobre* V′. *Então* T *é unitariamente equivalente a* T′ *se, e somente se,* T *e* T′ *possuem o mesmo polinômio característico.*

Demonstração. Suponhamos que T e T' possuam o mesmo polinômio característico f. Sejam $W_j \, (1 \leq j \leq k)$ as componentes primárias de V sob T e T_j a restrição de T a W_j. Suponhamos que I_j seja o operador idêntico sobre W_j. Então

$$f = \prod_{j=1}^{k} \det (xI_j - T_j).$$

Seja p_j o polinômio minimal de T_j. Se $p_j = X - c_j$ é claro que

$$\det (xI_j - T_j) = (x - c_j)^{s_j}$$

onde s_j é a dimensão de W_j. Por outro lado, se $p_j = (x - a_j)^2 + b_j^2$ com a_j e b_j reais e $b_j \neq 0$, então decorre do Teorema 18 que

$$\det (xI_j - T_j) = p_j^{s_j}$$

onde, nesse caso, $2s_j$ é a dimensão de W_j. Portanto, $f = \prod_j p_j^{s_j}$. Agora podemos também calcular f pelo mesmo método, usando as componentes primárias de V' sob T'. Como $p_1, \ldots, p_k$ são primos distintos, segue-se da unicidade da decomposição em fatores primos de f, que existem exatamente k componentes primárias $W'_j (1 \leq j \leq k)$ de V' sob T' e que elas podem ser indexadas de tal modo que p_j seja o polinômio minimal da restrição T'_j de T' a W'_j. Se $p_j = x - c_j$, então $T_j = c_j I_j$ e $T'_j = c_j I'_j$ onde I'_j é o operador idêntico sobre W'_j. Nesse caso é evidente que T_j é unitariamente equivalente a T'_j. Se $p_j = (x - a_j)^2 + b_j^2$, como acima, então usando o lema e o Teorema 20, vemos novamente que T_j é unitariamente equivalente a T'_j. Assim, para cada j existem bases ortonormais $\mathscr{B}_j$ e $\mathscr{B}'_j$ de W_j e W'_j, respectivamente, tais que

$$[T_j]_{\mathscr{B}_j} = [T'_j]_{\mathscr{B}'_j}$$

Finalmente, seja U a transformação linear de V em V' que transforma cada $\mathscr{B}_j$ sobre $\mathscr{B}'_j$. Então U é uma transformação unitária de V sobre V' tal que $UTU^{-1} = T'$.

Capítulo 10
FORMAS BILINEARES

10.1 Formas Bilineares

Neste capítulo vamos tratar das formas bilineares sobre espaços vetoriais de dimensão finita. O leitor provavelmente observará uma semelhança entre uma parte da matéria e a discussão dos determinantes no Capítulo 5 e dos produtos internos e formas nos Capítulos 8 e 9. A relação entre formas bilineares e produtos internos é particularmente forte; no entanto, este capítulo não pressupõe nada da matéria do Capítulo 8 ou 9. O leitor que não tiver familiaridade com produtos internos provavelmente lucraria lendo a primeira parte do Capítulo 8 à medida que lesse a discussão de formas bilineares.

A primeira seção trata do espaço das formas bilineares sobre um espaço vetorial de dimensão n. A matriz de uma forma bilinear em relação a uma base ordenada é introduzida e é estabelecido o isomorfismo entre o espaço das formas e o espaço das $n \times n$ matrizes. Define-se o posto de uma forma bilinear e são introduzidas as formas bilineares não degeneradas. A segunda seção discute as formas bilineares simétricas e sua diagonalização. A terceira seção estuda as formas bilineares anti-simétricas. A quarta seção discute o grupo que conserva uma forma bilinear não-degenerada, com atenção especial prestada aos grupos ortogonais, os grupos pseudo-ortogonais e um grupo pseudo-ortogonal particular — o grupo de Lorentz.

Definição. *Seja* V *um espaço vetorial sobre o corpo* F. *Uma* **forma bilinear** *sobre* V *é uma função* f, *que associa a cada par ordenado de vetores* α, β *em* V *um escalar* $f(\alpha, \beta)$ *em* F, *e que satisfaz*

$$f(c\alpha_1 + \alpha_2, \beta) = cf(\alpha_1, \beta) + f(\alpha_2, \beta)$$

(10-1)

$$f(\alpha, c\beta_1 + \beta_2) = cf(\alpha, \beta_1) + f(\alpha, \beta_2).$$

Indicando por $V \times V$ o conjunto de todos os pares ordenados de vetores em V, esta definição pode ser reformulada como segue: Uma forma bilinear sobre V é uma função f de $V \times V$ em F que é linear como uma função de qualquer dos seus argumentos quando o outro é deixado fixo. A função nula de $V \times V$ em F é evidentemente uma forma bilinear. É também verdade que toda combinação linear de formas bilineares sobre V é ainda uma forma bilinear. Para demonstrar este fato, basta considerar combinações lineares do tipo $cf + g$, sendo f e g formas bilineares sobre V. A demonstração de que $cf + g$ satisfaz (10-1) é semelhante a muitas outras que fizemos e vamos omiti-la. Tudo isto pode ser resumido dizendo-se que o conjunto das formas bilineares sobre V é um subespaço do espaço das funções de $V \times V$ em F (Exemplo 3, Capítulo 2). Indicaremos o espaço das formas bilineares sobre V por $L(V, V, F)$.

Exemplo 1. Seja V um espaço vetorial sobre o corpo F e sejam L_1 e L_2 funcionais lineares sobre V. Definamos f por

$$f(\alpha, \beta) = L_1(\alpha)L_2(\beta).$$

Fixando β e considerando f como uma função de α, então temos simplesmente um múltiplo escalar do funcional linear L_1. Com α fixo, f é um múltiplo escalar de L_2. Assim, é evidente que f é uma forma bilinear sobre V.

Exemplo 2. Sejam m e n inteiros positivos e F um corpo. Seja V o espaço vetorial das $m \times n$ matrizes sobre F. Seja A uma $m \times m$ matriz fixa sobre F. Definamos

$$f_A(X, Y) = \text{traço}(X^t A Y).$$

Então f_A é uma forma bilinear sobre V. De fato, se X, Y e Z são $m \times n$ matrizes sobre F,

$$\begin{aligned} f_A(cX + Z, Y) &= \text{traço}[(cX + Z)^t A Y] \\ &= \text{traço}(cX^t A Y) + \text{traço}(Z^t A Y) \\ &= cf_A(X, Y) + f_A(Z, Y). \end{aligned}$$

Evidentemente, utilizamos o fato de que a operação transposta e a função traço são lineares. É ainda mais fácil mostrar que f_A é linear

como uma função do seu segundo argumento. No caso particular $n = 1$, a matriz X^tAY é 1×1, isto é, um escalar, e a forma bilinear é simplesmente

$$\begin{aligned} f_A(X, Y) &= X^tAY \\ &= \sum_i \sum_j A_{ij}x_iy_j. \end{aligned}$$

Mostraremos em breve que toda forma bilinear sobre o espaço das $m \times 1$ matrizes é desse tipo, isto é, é f_A para alguma $m \times m$ matriz A.

Exemplo 3. Seja F um corpo. Vamos determinar todas as formas bilineares sobre o espaço F^2. Suponhamos que f seja uma tal forma bilinear. Se $\alpha = (x_1\ x_2)$ e $\beta = (y_1, y_2)$ são vetores em F^2, então

$$\begin{aligned} f(\alpha, \beta) &= f(x_1\varepsilon_1 + x_2\varepsilon_2, \beta) \\ &= x_1 f(\varepsilon_1, \beta) + x_2 f(\varepsilon_2, \beta) \\ &= x_1 f(\varepsilon_1, y_1\varepsilon_1 + y_2\varepsilon_2) + x_2 f(\varepsilon_2, y_1\varepsilon_1 + y_2\varepsilon_2) \\ &= x_1y_1 f(\varepsilon, \varepsilon_1) + x_1y_2 f(\varepsilon_1, \varepsilon_2) + x_2y_1 f(\varepsilon_2, \varepsilon_1) + x_2y_2 f(\varepsilon_2, \varepsilon_2). \end{aligned}$$

Assim, f é completamente determinada pelos quatro escalares $A_{ij} = f(\varepsilon_i, \varepsilon_j)$ da seguinte maneira:

$$\begin{aligned} f(\alpha, \beta) &= A_{11}x_1y_1 + A_{12}x_1y_2 + A_{21}x_2y_1 + A_{22}x_2y_2 \\ &= \sum_{i,j} A_{ij}x_iy_j. \end{aligned}$$

Se X e Y são as matrizes das coordenadas de α e β e se A é a 2×2 matriz com elementos $A(i, j) = A_{ij} = f(\varepsilon_i, \varepsilon_j)$, então

(10-2) $$f(\alpha, \beta) = X^tAY.$$

Observemos no Exemplo 2 que se A é uma 2×2 matriz arbitrária sobre F, então (10-2) define uma forma bilinear sobre F^2. Vemos que as formas bilineares sobre F^2 são exatamente as obtidas por meio de uma 2×2 matriz como em (10-2).

A discussão no Exemplo 3 pode ser generalizada de modo a descrever todas as formas bilineares sobre um espaço vetorial de dimensão finita. Seja V um espaço vetorial de dimensão finita sobre

o corpo F e seja $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ uma base ordenada de V. Suponhamos que f seja uma forma bilinear sobre V. Se

$$\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n \quad \text{e} \quad \beta = y_1\alpha_1 + \ldots + y_n\alpha_n$$

são vetores em V, então

$$\begin{aligned} f(\alpha, \beta) &= f(\sum_i x_i\alpha_i, \beta) \\ &= \sum_i x_i f(\alpha_i, \beta) \\ &= \sum_i x_i f(\alpha_i, \sum_i y_j\alpha_j) \\ &= \sum_i \sum_j x_i y_j f(\alpha_i, \alpha_j). \end{aligned}$$

Se fizermos $A_{ij} = f(\alpha_i, \alpha_j)$, então

$$\begin{aligned} f(\alpha, \beta) &= \sum_i \sum_j A_{ij} x_i y_j \\ &= X^t A Y \end{aligned}$$

onde X e Y são as matrizes das coordenadas de α e β em relação à base ordenada $\mathscr{B}$. Assim, toda forma bilinear sobre V é do tipo

$$(10\text{-}3) \qquad f(\alpha, \beta) = [\alpha]_{\mathscr{B}}^t A [\beta]_{\mathscr{B}}$$

para alguma $n \times n$ matriz A sobre F. Reciprocamente, se temos uma $n \times n$ matriz arbitrária A, é fácil ver que (10-3) define uma forma bilinear f sobre V, tal que $A_{ij} = f(\alpha_i, \alpha_j)$.

Definição. *Seja* V *um espaço vetorial de dimensão finita e seja* $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ *uma base ordenada de* V. *Se* f *é uma forma bilinear sobre* V, *a* **matriz de f em relação à base ordenada** $\mathscr{B}$ *é a* $n \times n$ *matriz* A *com elementos* $A_{ij} = f(\alpha_i, \alpha_j)$. *Às vezes, indicaremos esta matriz por* $[f]_{\mathscr{B}}$.

Teorema 1. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo* F. *Para cada base ordenada* $\mathscr{B}$ *de* V, *a função que associa a cada forma bilinear sobre* V *sua matriz em relação à base ordenada* $\mathscr{B}$ *é um isomorfismo do espaço* L(V, V, F) *no espaço das* $n \times n$ *matrizes sobre o corpo* F.

Demonstração. Observamos acima que $f \to [f]_{\mathscr{B}}$ é uma correspondência bijetora entre o conjunto das formas bilineares sobre V e o conjunto de todas as $n \times n$ matrizes sobre F. Que isso é uma transformação linear é fácil de ver, pois

$$(cf + g)(\alpha_i, \alpha_j) = cf(\alpha_i, \alpha_j) + g(\alpha_i, \alpha_j)$$

para todos i e j. Isto diz simplesmente que

$$[cf + g]_{\mathscr{B}} = c[f]_{\mathscr{B}} + [g]_{\mathscr{B}}.$$

Corolário. *Se* $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ *é uma base ordenada de* V *e* $\mathscr{B}^* = \{L_1, \ldots, L_n\}$ *é a base dual de* V*, *então as* n^2 *formas bilineares*

$$f_{ij}(\alpha, \beta) = L_i(\alpha)L_j(\beta), \quad 1 \leq i \leq n, \quad 1 \leq j \leq n$$

formam uma base do espaço L(V, V, F). *Em particular, a dimensão de* L(V, V, F) *é* n^2.

Demonstração. A base dual $\{L_1, \ldots, L_n\}$ é definida essencialmente pelo fato de que $L_i(\alpha)$ é a i-ésima coordenada de α em relação à base ordenada $\mathscr{B}$ (para todo α em V). Ora, as funções f_{ij} definidas por

$$f_{ij}(\alpha, \beta) = L_i(\alpha)L_j(\beta)$$

são formas bilineares do tipo considerado no Exemplo 1. Se

$$\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n \quad \text{e} \quad \beta = y_1\alpha_1 + \ldots + y_n\alpha_n,$$

então

$$f_{ij}(\alpha, \beta) = x_i y_j.$$

Seja f uma forma bilinear arbitrária sobre V e seja A a matriz de f em relação à base ordenada $\mathscr{B}$. Então

$$f(\alpha, \beta) = \sum_{i,j} A_{ij} x_i y_j$$

o que diz simplesmente que

$$f = \sum_{i,j} A_{ij} f_{ij}.$$

Agora é evidente que as n^2 formas f_{ij} formam uma base de $L(V, V, F)$.

Pode-se reformular a demonstração do corolário como segue. A matriz da forma bilinear f_{ij} em relação à base ordenada $\mathscr{B}$ é a matriz "unitária" $E^{i,j}$, cujo único elemento não-nulo é um 1 na linha i e coluna j. Como estas matrizes $E^{i,j}$ constituem uma base do espaço das $n \times n$ matrizes, as formas f_{ij} constituem uma base do espaço das formas bilineares.

O conceito de matriz de uma forma bilinear em relação a uma base ordenada é semelhante ao conceito de matriz de um operador linear em relação a uma base ordenada. Do mesmo modo que para operadores lineares, estaremos interessados no que acontece à matriz que representa uma forma bilinear, ao passarmos de uma base ordenada a outra. Então, suponhamos que $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ e $\mathscr{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ sejam duas bases ordenadas de V e que f seja uma forma bilinear sobre V. Como se relacionam as matrizes $[f]_{\mathscr{B}}$ e $[f]_{\mathscr{B}'}$? Bem, seja P a $n \times n$ matriz (inversível) tal que

$$[\alpha]_{\mathscr{B}} = P[\alpha]_{\mathscr{B}'}$$

para todo α em V. Em outras palavras, definimos P por

$$\alpha'_j = \sum_{i=1}^{n} P_{ij}\alpha_i.$$

Para vetores arbitrários α, β em V

$$\begin{aligned} f(\alpha, \beta) &= [\alpha]^t_{\mathscr{B}}[f]_{\mathscr{B}}[\beta]_{\mathscr{B}} \\ &= (P[\alpha]_{\mathscr{B}'})^t[f]_{\mathscr{B}}P[\beta]_{\mathscr{B}'} \\ &= [\alpha]^t_{\mathscr{B}'}(P^t[f]_{\mathscr{B}}P)[\beta]_{\mathscr{B}'}. \end{aligned}$$

Pela definição e unicidade da matriz que representa f em relação à base ordenada $\mathscr{B}'$, devemos ter

$$[f]_{\mathscr{B}'} = P^t[f]_{\mathscr{B}}P. \tag{10-4}$$

Exemplo 4. Seja V o espaço vetorial R^2. Seja f a forma bilinear definida sobre $\alpha = (x_1, x_2)$ e $\beta = (y_1, y_2)$ por

$$f(\alpha, \beta) = x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2.$$

Ora,

$$f(\alpha, \beta) = [x_1, x_2] \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

e, então, a matriz de f em relação à base ordenada canônica $\mathscr{B} = \{\varepsilon_1, \varepsilon_2\}$ é

$$[f]_{\mathscr{B}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Seja $\mathscr{B}' = \{\varepsilon_1', \varepsilon_2'\}$ a base ordenada definida por $\varepsilon_1' = (1, -1)$, $\varepsilon_2' = (1, 1)$. Neste caso, a matriz P que transforma as coordenadas de $\mathscr{B}'$ para $\mathscr{B}$ é

$$P = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Assim,

$$\begin{aligned} [f]_{\mathscr{B}'} &= P^t [f]_{\mathscr{B}} P \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}. \end{aligned}$$

O que isto significa é que, se exprimirmos os vetores α e β por meio de suas coordenadas em relação à base $\mathscr{B}'$, digamos

$$\alpha = x_1'\varepsilon_1' + x_2'\varepsilon_2', \quad \beta = y_1'\varepsilon_1' + y_2'\varepsilon_2'$$

então

$$f(\alpha, \beta) = 4x_2'y_2'.$$

Uma conseqüência da fórmula (10-4) da mudança de base é a seguinte: Se A e B são $n \times n$ matrizes que representam a mesma forma bilinear sobre V em relação a bases ordenadas (possivelmente) diferentes, então A e B têm o mesmo posto. De fato, se P

é uma $n \times n$ matriz inversível e $B = P^t AP$, é evidente que A e B têm o mesmo posto. Isto torna possível definir o posto de uma **forma bilinear** *sobre* V *é uma função* f, *que associa a cada par orde-* que represente a forma em relação a uma base ordenada de V.

É desejável dar uma definição mais intrínseca do posto de uma forma bilinear. Isto pode ser feito como segue: Suponhamos que f seja uma forma bilinear sobre o espaço vetorial V. Fixando um vetor α em V, $f(\alpha, \beta)$ é linear com uma função de β. Desta maneira, cada α fixo determina um funcional linear sobre V; indiquemos este funcional linear por $L_f(\alpha)$. Repetindo, se α é um vetor em V, então $L_f(\alpha)$ é o funcional linear sobre V cujo valor em qualquer vetor β é $f(\alpha, \beta)$. Isto nos dá uma transformação $\alpha \rightarrow L_f(\alpha)$ de V no espaço dual V^*. Como

$$f(c\alpha_1 + \alpha_2, \beta) = cf(\alpha_1, \beta) + f(\alpha_2, \beta)$$

vemos que

$$L_f(c\alpha_1 + \alpha_2) = cL_f(\alpha_1) + L_f(\alpha_2)$$

isto é, L_f é uma transformação linear de V em V^*.

De maneira semelhante, f determina uma transformação linear R_f de V em V^*. Para cada β fixo em V, $f(\alpha, \beta)$ é linear como uma função de α. Definimos $R_f(\beta)$ como sendo o funcional linear sobre V cujo valor no vetor α é $f(\alpha, \beta)$.

Teorema 2. *Seja* f *uma forma bilinear sobre o espaço vetorial* V *de dimensão finita. Sejam* L_f *e* R_f *as transformações lineares de* V *em* V* *definidas por* $(L_f\alpha)(\beta) = f(\alpha, \beta) = (R_f\beta)(\alpha)$. *Então, posto* $(L_f) =$ *posto* (R_f).

Demonstração. Pode-se fazer uma demonstração deste teorema que seja "independente de coordenadas". Tal demonstração é semelhante à demonstração (na Seção 3.7) de que o posto-linha de uma matriz é igual ao seu posto-coluna. Então, faremos aqui uma demonstração que começa tomando um sistema de coordenadas (base), utilizando depois o teorema "posto-linha igual a posto-coluna".

Para demonstrar que posto $(L_f) =$ posto (R_f) bastará demonstrar que L_f e R_f têm a mesma nulidade. Seja $\mathcal{B}$ uma base ordenada de V e seja $A = [f]_{\mathcal{B}}$. Se α e β são vetores em V, com matrizes de coordenadas X e Y em relação à base ordenada $\mathcal{B}$, então

$f(\alpha, \beta) = X^t A Y$. Ora, $R_f(\beta) = 0$ significa que $f(\alpha, \beta) = 0$ para todo α em V, isto é, que $X^t A Y = 0$ para toda $n \times 1$ matriz X. A última condição diz simplesmente que $AY = 0$. A nulidade de R_f é, portanto, igual à dimensão do espaço das soluções de $AY = 0$.

Analogamente, $L_f(\alpha) = 0$ se, e somente se, $X^t A Y = 0$ para toda $n \times 1$ matriz Y. Assim, α está no núcleo de L_f se, e somente se, $X^t A = 0$, isto é, $A^t X = 0$. A nulidade de L_f é, portanto, igual à dimensão do espaço das soluções de $A^t X = 0$. Como as matrizes A e A^t têm o mesmo posto-coluna, vemos que

$$\text{nulidade}\,(L_f) = \text{nulidade}\,(R_f).$$

Definição. *Se* f *é uma forma bilinear sobre o espaço* V *de dimensão finita, o* **posto** *de* f *é o inteiro* $r = posto\ (L_f) = posto\ (R_f)$.

Corolário 1. *O posto de uma forma bilinear é igual ao posto da matriz da forma bilinear em relação a qualquer base ordenada.*

Corolário 2. *Se* f *é uma forma bilinear sobre o espaço vetorial* n-*dimensional* V, *as seguintes afirmações são equivalentes*:

(a) *Posto* (f) = n.

(b) *Para cada* α *não-nulo em* V, *existe um* β *em* V *tal que* $f(\alpha, \beta) \neq 0$.

(c) *Para cada* β *não-nulo em* V, *existe um* α *em* V *tal que* $f(\alpha, \beta) \neq 0$.

Demonstração. A afirmação (b) diz simplesmente que o núcleo de L_f é o subespaço nulo. A afirmação (c) diz que o núcleo de R_f é o subespaço nulo. As transformações lineares L_f e R_f têm nulidade 0 se, e somente se, elas têm o mesmo posto n, isto é, se, e somente se, posto $(f) = n$.

Definição. *Uma forma bilinear* f *sobre um espaço vetorial* V *é dita* **não-degenerada** (*ou* **não-singular**) *se satisfaz as condições* (b) *e* (c) *do Corolário* 2.

Se V é de dimensão finita, então f é não-degenerada, desde que f satisfaça qualquer uma das três condições do Corolário 2. Em particular, f é não-degenerada (não-singular) se, e somente se, sua matriz em relação a alguma (toda) base ordenada de V é uma matriz não-singular.

Exemplo 5. Seja $V = R^n$ e seja f a forma bilinear definida sobre $\alpha = (x_1, \ldots, x_n)$ e $\beta = (y_1, \ldots, y_n)$ por

$$f(\alpha, \beta) = x_1 y_1 + \ldots + x_n y_n.$$

Então f é uma forma bilinear não-degenerada sobre R^n. A matriz de f em relação à base ordenada canônica é a $n \times n$ matriz unidade:

$$f(X, Y) = X^t Y.$$

Esta f é usualmente denominada o produto escalar. O leitor provavelmente tem familiaridade com esta forma bilinear, pelo menos no caso $n = 3$. Geometricamente, o número $f(\alpha, \beta)$ é o produto do comprimento de α pelo comprimento de β e pelo cosseno do ângulo entre α e β. Em particular, $f(\alpha, \beta) = 0$ se, e somente se, os vetores α e β são ortogonais (perpendiculares).

Exercícios

1. Quais das seguintes funções f, definidas sobre vetores $\alpha = (x_1, x_2)$ e $\beta = (y_1\ y_2)$ em R^2, são formas bilineares?

(a) $f(\alpha, \beta) = 1$;

(b) $f(\alpha, \beta) = (x_1 - y_1)^2 + x_2 y_2$;

(c) $f(\alpha, \beta) = (x_1 + y_1)^2 - (x_1 - y_1)^2$;

(d) $f(\alpha, \beta) = x_1 y_2 - x_2 y_1$.

2. Seja f a forma bilinear sobre R^2 definida por

$$f((x_1, y_1), (x_2, y_2)) = x_1 y_1 + x_2 y_2.$$

Determinar a matriz de f em relação a cada uma das seguintes bases:

$$\{(1, 0), (0, 1)\}, \quad \{(1, -1), (1, 1)\}, \quad \{(1, 2), (3, 4)\}.$$

3. Seja V o espaço das 2×3 matrizes sobre R e seja f a forma bilinear sobre V definida por $f(X, Y) = \text{traço}\ (X^t A Y)$, onde

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Determinar a matriz de f em relação à base ordenada

$$\{E^{11}, E^{12}, E^{13}, E^{21}, E^{22}, E^{23}\}$$

onde E^{ij} é a matriz cujo único elemento não-nulo é um 1 na linha i e coluna j.

4. Descrever explicitamente todas as formas bilineares f sobre R^3 com a propriedade de que $f(\alpha, \beta) = f(\beta, \alpha)$ para todos α, β.

5. Descrever as formas bilineares sobre R^3 que satisfazem $f(\alpha, \beta) = -f(\beta, \alpha)$ para todos α, β.

6. Seja n um inteiro positivo e seja V o espaço das $n \times n$ matrizes sobre o corpo dos números complexos. Mostrar que a equação

$$f(A, B) = n\,\mathrm{tr}(AB) - \mathrm{tr}(A)\,\mathrm{tr}(B)$$

define uma forma bilinear f sobre V. É verdade que $f(A, B) = f(B, A)$ para todas A, B?

7. Seja f a forma bilinear do Exercício 6. Mostrar que f é degenerada (f não é não-degenerada). Seja V_1 o subespaço de V formado pelas matrizes de traço 0 e seja f_1 a restrição de f a V_1. Mostrar que f_1 é não-degenerada.

8. Seja f a forma bilinear definida no Exercício 6 e seja V_2 o subespaço de V formado por todas as matrizes A tais que traço $(A) = 0$ e $A^* = -A$. (A^* é a transposta conjugada de A.) Indiquemos por f_2 a restrição de f a V_2. Mostrar que f_2 é negativa definida, isto é, que $f_2(A, A) < 0$ para toda A não-nula em V_2.

9. Seja f a forma bilinear definida no Exercício 6. Seja W o conjunto das matrizes A em V tais que $f(A, B) = 0$ para toda B. Mostrar que W é um subespaço de V. Descrever W explicitamente e determinar sua dimensão.

10. Seja f uma forma bilinear sobre um espaço vetorial V de dimensão finita. Seja W o subespaço formado pelos β tais que $f(\alpha, \beta) = 0$ para todo α. Mostrar que

$$\text{posto } (f) = \dim V - \dim W.$$

Usar este resultado e o resultado do Exercício 9 para calcular o posto da forma bilinear definida no Exercício 6.

11. Seja f uma forma bilinear sobre um espaço vetorial V de dimensão finita. Suponhamos que V_1 seja um subespaço de V com a propriedade de que a restrição de f a V_1 seja não-degenerada. Mostrar que posto $(f) \geq \dim V_1$.

12. Sejam f e g formas bilineares sobre um espaço vetorial V de dimensão finita. Suponhamos que g seja não-singular. Mostrar que existem operadores lineares T_1, T_2 sobre V_1 únicos, tais que

$$f(\alpha, \beta) = g(T_1\alpha, \beta) = g(\alpha, T_2\beta)$$

para todos α, β.

13. Mostrar que o resultado do Exercício 12 não é necessariamente válido se g é singular.

14. Seja f uma forma bilinear sobre um espaço vetorial V de dimensão finita. Mostrar que f pode ser expresso como um produto de dois funcionais lineares (isto é, $f(\alpha\ \beta) = L_1(\alpha)\ L_2(\beta)$ para L_1, L_2 em V^*) se, e somente se, f tem posto 1.

10.2 Formas Bilineares Simétricas

O propósito principal desta seção é responder à seguinte pergunta: Se f é uma forma bilinear sobre o espaço vetorial V de dimensão finita, quando é que existe uma base ordenada $\mathscr{B}$ de V, em relação à qual f é representada por uma matriz diagonal? Vamos demonstrar que isto é possível se, e somente se, f é uma forma bilinear simétrica, ou seja, $f(\alpha, \beta) = f(\beta, \alpha)$. O teorema será demonstrado apenas para o caso do corpo de escalares ser um corpo F de característica zero, isto é, que, se n é um inteiro positivo, a soma $1 + \ldots + 1$ (n vezes) em F não é 0.

Definição. *Seja* f *uma forma bilinear sobre o espaço vetorial* V. *Dizemos que* f *é* **simétrica** *se* $f(\alpha, \beta) = f(\beta, \alpha)$ *para todos os vetores* α, β *em* V.

Se V é de dimensão finita, a forma bilinear f é simétrica se, e somente se, sua matriz A em relação a alguma ou (toda) base ordenada é simétrica, isto é, $A^t = A$. Para ver isto, perguntamos quando é que a forma bilinear

$$f(X, Y) = X^t A Y$$

é simétrica. Isto acontece se, e somente se, $X^t A Y = Y^t A X$ para todas matrizes-colunas X e Y. Como $X^t A Y$ é uma 1×1 matriz, temos $X^t A Y = Y^t A^t X$. Assim, f é simétrica se, e somente se, $Y^t A^t X = Y^t A X$ para todas X, Y. Evidentemente, isto significa apenas que $A = A^t$. Em particular, deve-se notar que se existir uma base ordenada de V em relação à qual f seja representada por uma matriz diagonal, então f é simétrica, pois qualquer matriz diagonal é uma matriz simétrica.

Se f é uma forma bilinear simétrica, a **forma quadrática associada a** f é a função q de V em F definida por

$$q(\alpha) = f(\alpha, \alpha).$$

Se F é um subcorpo do corpo dos números complexos, a forma bilinear simétrica f é completamente determinada por sua forma quadrática associada, de acordo com a **identidade de polarização.**

$$(10\text{-}5)\qquad f(\alpha, \beta) = \tfrac{1}{4}q(\alpha + \beta) - \tfrac{1}{4}q(\alpha - \beta).$$

A demonstração de (10-5) requer apenas cálculos de rotina, que omitiremos. Se f é a forma bilinear do Exemplo 5, ou seja, o produto escalar, então a forma quadrática associada é

$$q(x_1, \ldots, x_n) = x_1^2 + \ldots + x_n^2.$$

Em outras palavras, $q(\alpha)$ é o quadrado do comprimento de α. Para a forma bilinear $f_A(X, Y) = X^tAY$, a forma quadrática associada é

$$q_A(X) = X^tAX = \sum_{i,j} A_{ij}x_ix_j.$$

Uma classe importante de formas bilineares simétricas consiste dos produtos internos sobre espaços vetoriais reais, discutidos no Capítulo 8. Se V é um espaço vetorial *real*, um **produto interno** sobre V é uma forma bilinear simétrica f sobre V que satisfaz

$$(10\text{-}6)\qquad f(\alpha, \alpha) > 0 \quad \text{se} \quad \alpha \neq 0.$$

Uma forma bilinear que satisfaz (10-6) é dita **positiva definida**. Assim, um produto interno sobre um espaço vetorial real é uma forma bilinear simétrica positiva definida sobre aquele espaço. Notemos que um produto interno é não-degenerado. Dois vetores α, β são ditos **ortogonais** em relação ao produto interno f se $f(\alpha, \beta) = 0$. A forma quadrática $q(\alpha) = f(\alpha, \alpha)$ toma apenas valores não-negativos e $q(\alpha)$ é usualmente considerado como sendo o quadrado do comprimento de α. Evidentemente, estes conceitos de comprimento e ortogonalidade se originam do exemplo mais importante de produto interno — o produto escalar do Exemplo 5.

Se f é uma forma bilinear simétrica sobre um espaço vetorial V, é conveniente aplicar um pouco da terminologia de produtos internos a f. É particularmente conveniente dizer que α e β são ortogonais em relação a f se $f(\alpha, \beta) = 0$. Não é aconselhável considerar $f(\alpha, \alpha)$ como sendo o quadrado do comprimento de α; por exemplo, se V é um espaço vetorial complexo, podemos ter $f(\alpha, \alpha) = \sqrt{-1}$, ou num espaço vetorial real $f(\alpha, \alpha) = -2$.

Passamos, agora, ao teorema fundamental desta seção. Ao ler a demonstração, o leitor deverá achar útil pensar no caso particular em que V é um espaço vetorial real e f é um produto interno sobre V.

Teorema 3. *Seja* V *um espaço vetorial de dimensão finita sobre um corpo de característica zero e seja* f *uma forma bilinear simétrica sobre* V. *Então existe uma base ordenada de* V *em relação à qual* f *é representada por uma matriz diagonal.*

Demonstração. O que precisamos encontrar é uma base ordenada

$$\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$$

tal que $f(\alpha_i, \alpha_j) = 0$ para $i \neq j$. Se $f = 0$ ou $n = 1$, o teorema é obviamente verdadeiro. Assim, podemos supor $f \neq 0$ e $n > 1$. Se $f(\alpha, \alpha) = 0$ para todo α em V, a forma quadrática associada q é identicamente 0 e a identidade de polarização (10-5) mostra que $f = 0$. Assim, existe um vetor α em V tal que $f(\alpha, \alpha) = q(\alpha) \neq 0$. Seja W o subespaço unidimensional de V que é gerado por α e seja $W^\perp$ o conjunto dos vetores β em V tais que $f(\alpha, \beta) = 0$. Afirmamos, agora, que $V = W \oplus W^\perp$. Certamente os subespaços W e $W^\perp$ são independentes. Um vetor típico em W é $c\alpha$, onde c é um escalar. Se $c\alpha$ está, também, em $W^\perp$, então $f(c\alpha, c\alpha) = c^2 f(\alpha, \alpha) = 0$. Mas, $f(\alpha, \alpha) \neq 0$, logo, $c = 0$. Além disso, todo vetor em V é a soma de um vetor em W e um vetor em $W^\perp$. De fato, seja γ um vetor arbitrário em V e coloquemos

$$\beta = \gamma - \frac{f(\gamma, \alpha)}{f(\alpha, \alpha)}\alpha.$$

Então

$$f(\alpha, \beta) = f(\alpha, \gamma) - \frac{f(\gamma, \alpha)}{f(\alpha, \alpha)} f(\alpha, \alpha)$$

e, como f é simétrica, $f(\alpha, \beta) = 0$. Portanto, β está no subespaço $W^\perp$. A expressão

$$\gamma = \frac{f(\gamma, \alpha)}{f(\alpha, \alpha)}\alpha + \beta$$

nos mostra que $V = W + W^\perp$.

A restrição de f a $W^\perp$ é uma forma bilinear simétrica sobre $W^\perp$. Como $W^\perp$ tem dimensão $(n - 1)$, podemos supor, por indução, que $W^\perp$ possua uma base $\{\alpha_2, \ldots, \alpha_n\}$ tal que

$$f(\alpha_i, \alpha_j) = 0, \quad i \neq j \quad (i \geq 2, j \geq 2).$$

Colocando $\alpha_1 = \alpha$, obtemos uma base $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que $f(\alpha_i, \alpha_j) = 0$ para $i \neq j$.

Corolário. *Seja* F *um subcorpo do corpo dos números complexos e seja* A *uma* n × n *matriz simétrica sobre* F. *Então existe um* n × n *matriz inversível* P *sobre* F *tal que* P^tAP *seja diagonal.*

No caso de F ser o corpo dos números reais, a matriz inversível P neste corolário pode ser escolhida de modo a ser uma matriz *ortogonal*, isto é, $P^t = P^{-1}$. Em outras palavras, se A é uma $n \times n$ matriz simétrica real, existe uma matriz ortogonal real P tal que P^tAP seja diagonal; contudo, isto não é nada evidente a partir do que fizemos acima (ver Capítulo 8).

Teorema 4. *Seja* V *um espaço vetorial de dimensão finita sobre o corpo dos números complexos. Seja* f *uma forma bilinear simétrica sobre* V *que tenha posto* r. *Então existe uma base ordenada* $\mathscr{B} = \{\beta_1, \ldots, \beta_n\}$ *de* V *tal que*

(i) *a matriz de* f *em relação à base ordenada* $\mathscr{B}$ *é diagonal*;

(ii) $f(\beta_j, \beta_j) = \begin{cases} 1, & j = 1, \ldots, r \\ 0, & j > r. \end{cases}$

Demonstração. Pelo Teorema 3, existe uma base ordenada $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que

$$f(\alpha_i, \alpha_j) = 0 \quad \text{para} \quad i \neq j.$$

Como f tem posto r, sua matriz em relação à base ordenada $\{\alpha_1, \ldots, \alpha_n\}$ também o tem. Assim, precisamos ter $f(\alpha_j, \alpha_j) \neq 0$ para exatamente r valores de j. Reordenando os vetores α_j, podemos admitir que

$$f(\alpha_j, \alpha_j) \neq 0, \quad j = 1, \ldots, r.$$

Usemos, agora, o fato de que o corpo de escalares é o corpo dos números complexos. Se $\sqrt{f(\alpha_j, \alpha_j)}$ indica uma raiz quadrada complexa qualquer de $f(\alpha_j, \alpha_j)$ e se colocarmos

$$\beta_j = \begin{cases} \dfrac{1}{\sqrt{f(\alpha_j, \alpha_j)}}\,\alpha_j, & j = 1, \ldots, r \\[2ex] \alpha_j, & j > r \end{cases}$$

a base $\{\beta_1, \ldots, \beta_n\}$ satisfará as condições (i) e (ii).

Evidentemente, o Teorema 4 é válido se o corpo de escalares é um subcorpo qualquer do corpo dos números complexos, no qual todo elemento possua uma raiz quadrada. Não é válido, por exemplo, quando o corpo de escalares é o corpo dos números reais. Sobre o corpo dos números reais, temos o seguinte substituto para o Teorema 4.

Teorema 5. *Seja* V *um espaço vetorial* n*-dimensional sobre o corpo dos números reais e seja* f *uma forma bilinear simétrica sobre* V *que tenha posto* r. *Então existe uma base ordenada* $\{\beta_1, \ldots, \beta_n\}$ *de* V *em relação à qual a matriz de* f *é diagonal e tal que*

$$f(\beta_j, \beta_j) = \pm 1, \quad j = 1, \ldots, r.$$

Além disso, o número de vetores β_j *da base para os quais* $f(\beta_j, \beta_j) = 1$ *é independente da escolha da base.*

Demonstração. Existe uma base $\{\alpha_1, \ldots, \alpha_n\}$ de V tal que

$$\begin{aligned} f(\alpha_i, \alpha_j) &= 0, & i &\neq j. \\ f(\alpha_j, \alpha_j) &\neq 0, & 1 &\leq j \leq r \\ f(\alpha_j, \alpha_j) &= 0, & j &> r. \end{aligned}$$

Seja

$$\begin{aligned} \beta_j &= |f(\alpha_j, \alpha_j)|^{-1/2}\alpha_j, & 1 &\leq j \leq r \\ \beta_j &= \alpha_j, & j &> r \end{aligned}$$

Então $\{\beta_1, \ldots, \beta_n\}$ é uma base que tem as propriedades enunciadas.

Seja p o número de vetores β_j da base para os quais $f(\beta_j, \beta_j) = 1$: precisamos mostrar que o número p é independente da particular base que tomemos satisfazendo as condições acima. Seja V^+ o subespaço de V gerado pelos vetores β_j da base para os quais $f(\beta_j, \beta_j) = 1$, e seja V^- o subespaço gerado pelos vetores β_j da base tais que $f(\beta_j, \beta_j) = -1$. Ora, $p = \dim V^+$, logo o que precisamos demonstrar é a unicidade da dimensão de V^+. É fácil ver que se α é um vetor não-nulo em V^+, então $f(\alpha, \alpha) > 0$; em outras palavras, f é positiva definida sobre o subespaço V^+. Analogamente, se α é um vetor não-nulo em V^-, então $f(\alpha, \alpha) < 0$, isto é, f é negativa definida sobre o subespaço V^-. Seja, agora, $V^\perp$ o subespaço gerado pelos vetores β_j da base para os quais $f(\beta_j, \beta_j) = 0$. Se α está em $V^\perp$, então $f(\alpha, \beta) = 0$ para todo β em V.

Como $\{\beta_1, \ldots, \beta_n\}$ é uma base de V, temos

$$V = V^+ \oplus V^- \oplus V^\perp.$$

Além disso, afirmamos que se W é um subespaço arbitrário de V sobre o qual f seja positiva definida, então os subespaços W, V^- e $V^\perp$ são independentes. De fato, suponhamos que α esteja em W, β em V^-, γ em $V^\perp$ e que $\alpha + \beta + \gamma = 0$. Então

$$0 = f(\alpha, \alpha + \beta + \gamma) = f(\alpha, \alpha) + f(\alpha, \beta) + f(\alpha, \gamma)$$
$$0 = f(\beta, \alpha + \beta + \gamma) = f(\beta, \alpha) + f(\beta, \beta) + f(\beta, \gamma).$$

Como γ está em $V^\perp$, $f(\alpha, \gamma) = f(\beta, \gamma) = 0$; como f é simétrica, obtemos

$$0 = f(\alpha, \alpha) + f(\alpha, \beta)$$
$$0 = f(\beta, \beta) + f(\alpha, \beta)$$

logo $f(\alpha, \alpha) = f(\beta, \beta)$. Como $f(\alpha, \alpha) \geq 0$ e $f(\beta, \beta) \leq 0$, segue que

$$f(\alpha, \alpha) = f(\beta, \beta) = 0.$$

Mas f é positiva definida sobre W e negativa definida sobre V^-. Concluímos que $\alpha = \beta = 0$, e portanto que $\gamma = 0$ também.

Como

$$V = V^+ \oplus V^- \oplus V^\perp$$

e W, V^-, $V^\perp$ são independentes, vemos que $\dim W \leq \dim V^+$. Isto é, se W é um subespaço arbitrário de V sobre o qual f é positiva definida, a dimensão de W não pode exceder a dimensão de V^+. Se $\mathscr{B}_1$ é uma outra base ordenada de V que satisfaz as condições do teorema, teremos subespaços correspondentes V_1^+, V_1^- e $V_1^\perp$; o argumento acima mostra que $\dim V_1^+ \leq \dim V^+$. Invertendo o argumento, obtemos $\dim V^+ \leq \dim V_1^+$ e, conseqüentemente,

$$\dim V^+ = \dim V_1^+.$$

Existem diversos comentários que devem ser feitos acerca da base $\{\beta_1, \ldots, \beta_n\}$ do Teorema 5 e dos subespaços associados V^+, V^- e $V^\perp$. Primeiro, notemos que $V^\perp$ é exatamente o subes-

paço dos vetores que são "ortogonais" a todo espaço V. Observamos acima que $V^{\perp}$ está contido neste subespaço; mas,

$$\dim V^{\perp} = \dim V - (\dim V^{+} + \dim V^{-}) = \dim V - \text{posto}(f)$$

portanto, todo vetor α tal que $f(\alpha, \beta) = 0$ para todo β deve estar em $V^{\perp}$. Assim, o subespaço $V^{\perp}$ é único. Os subespaços V^{+} e V^{-} não são únicos; contudo, suas dimensões são únicas. A demonstração do Teorema 5 nos mostra que $\dim V^{+}$ é a máxima dimensão possível para qualquer subespaço sobre o qual f seja positiva definida. Analogamente, $\dim V^{-}$ é a máxima dimensão de qualquer subespaço sobre o qual f seja negativa definida. É claro que

$$\dim V^{+} + \dim V^{-} = \text{posto}(f).$$

O número

$$\dim V^{+} - \dim V^{-}$$

freqüentemente é denominado a **assinatura** de f. Ela é introduzida porque as dimensões de V^{+} e V^{-} são facilmente determinadas a partir do posto de f e da assinatura de f.

Talvez devamos fazer um comentário final a respeito da relação entre formas bilineares simétricas sobre espaços vetoriais reais e produtos internos. Suponhamos que V seja um espaço vetorial real de dimensão finita e que V_1, V_2, V_3 sejam subespaços de V tais que

$$V = V_1 \oplus V_2 \oplus V_3.$$

Suponhamos que f_1 seja um produto interno sobre V_1 e f_2 seja um produto interno sobre V_2. Podemos, então, definir uma forma bilinear simétrica f sobre V como segue: Se α, β são vetores em V, então podemos escrever

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3 \quad \text{e} \quad \beta = \beta_1 + \beta_2 + \beta_3$$

com α_j e β_j em V_j. Seja

$$f(\alpha, \beta) = f_1(\alpha_1, \beta_1) - f_2(\alpha_2, \beta_2).$$

O subespaço $V^{\perp}$ para f será V_3, V_1 é um V^{+} conveniente para f e V_2 é um V^{-} conveniente. Uma parte do enunciado do Teorema 5

é que toda forma bilinear simétrica sobre V surge desta maneira. O conteúdo adicional do teorema é que um produto interno é representado em relação a alguma base ordenada pela matriz unidade.

Exercícios

1. As seguintes expressões definem formas quadráticas q sobre R^2. Determinar a forma bilinear simétrica f correspondente a cada q,

   (a) ax_1^2.

   (b) bx_1x_2.

   (c) cx_2^2.

   (d) $2x_1^2 - \frac{1}{3}x_1x_2$.

   (e) $x_1^2 + 9x_2^2$.

   (f) $3x_1x_2 - x_2^2$.

   (g) $4x_1^2 + 6x_1x_2 - 3x_2^2$.

2. Determinar a matriz, em relação à base ordenada canônica, e o posto de cada uma das formas bilineares do Exercício 1. Indicar quais formas são não-degeneradas.

3. Seja $q(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2$ a forma quadrática associada a uma forma bilinear simétrica f sobre R^2. Mostrar que f é não-degenerada se, e somente se, $b^2 - 4ac \neq 0$.

4. Seja V um espaço vetorial de dimensão finita sobre um subcorpo F do corpo dos números complexos e seja S o conjunto das formas bilineares simétricas sobre V.

   (a) Mostrar que S é um subespaço de $L(V, V, F)$.

   (b) Determinar dim S.

   Seja Q o conjunto de todas as formas quadráticas sobre V.

   (c) Mostrar que Q é um subespaço do espaço de todas as funções de V em F.

   (d) Descrever explicitamente um isomorfismo T de Q em S, sem referência a qualquer base.

   (e) Seja U um operador linear sobre V e q um elemento de Q. Mostrar que a equação $(U^+q)(\alpha) = q(U\alpha)$ define uma forma quadrática U^+q sobre V.

   (*f*) Se U é um operador linear sobre V, mostrar que a função U^+ definida na parte (e) é um operador linear sobre Q. Mostrar que U^+ é inversível se, e somente se, U é inversível.

5. Seja q a forma quadrática sobre R^2 dada por

$$q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2,\ a \neq 0.$$

   Determinar um operador linear inversível U sobre R^2 tal que

$$(U^+q)(x_1, x_2) = ax_1^2 + \left(c - \frac{b^2}{a}\right)x_2^2.$$

(*Sugestão*: Para determinar U^{-1} (e portanto U), completar o quadrado. Para a definição de U^{+}, ver a parte (e) do Exercício 4.)

6. Seja q a forma quadrática sobre R^2 dada por

$$q(x_1, x_2) = 2bx_1x_2.$$

Determinar um operador linear inversível U sobre R^2 tal que

$$(U^{+}q)(x_1, x_2) = 2bx_1^2 - 2bx_2^2.$$

7. Seja q a forma quadrática sobre R^3 dada por

$$q(x_1, x_2, x_3) = x_1x_2 + 2x_1x_3 + x_3^2.$$

Determinar um operador linear inversível U sobre R^3 tal que

$$(U^{+}q)(x_1, x_2, x_3) = x_1^2 - x_2^2 + x_3^2.$$

(*Sugestão*: Exprimir U como um produto de operadores semelhantes àqueles usados nos Exercícios 5 e 6.)

8. Seja A uma $n \times n$ matriz simétrica sobre R e seja q a forma quadrática sobre R^n dada por

$$q(x_1, \ldots, x_n) = \sum_{i,j} A_{ij}x_ix_j.$$

Generalizar o método usado no Exercício 7 para mostrar que existe um operador linear inversível U sobre R^n tal que

$$(U^{+}q)(x_1, \ldots, x_n) = \sum_{i=1}^{n} c_ix_i^2$$

onde c_i é 1, -1 ou 0, $i = 1, \ldots, n$.

9. Seja f uma forma bilinear simétrica sobre R^n. Usar o resultado do Exercício 8 para demonstrar a existência de uma base ordenada $\mathcal{B}$ tal que $[f]_{\mathcal{B}}$ seja diagonal.

0. Seja V o espaço vetorial real das 2×2 matrizes hermitianas (complexas), isto é, 2×2 matrizes complexas A que satisfazem $A_{ij} = \overline{A_{ji}}$.

(a) Mostrar que a equação $q(A) = \det A$ define uma forma quadrática q sobre V.

(b) Seja W o subespaço de V formado pelas matrizes de traço 0. Mostrar que a forma bilinear f determinada por q é negativa definida sobre o subespaço W.

1. Seja V um espaço vetorial de dimensão finita e f uma forma bilinear simétrica não-degenerada sobre V. Mostrar que para cada operador linear T

sobre V existe um único operador T' sobre V tal que $f(T\alpha, \beta) = f(\alpha, T'\beta)$ para todos α, β em V. Mostrar também que

$$(T_1 T_2)' = T_2' T_1'$$
$$(c_1 T_1 + c_2 T_2)' = c_1 T_1' + c_2 T_2$$
$$(T')' = T.$$

Quanto disto acima continua válido sem a hipótese de que f é não-degenerada?

12. Seja F um corpo e V o espaço das $n \times 1$ matrizes sobre F. Suponhamos que A seja uma $n \times n$ matriz fixa sobre F e f seja a forma bilinear sobre V definida por $f(X, Y) = X^t A Y$. Suponhamos que f seja simétrica e não-degenerada. Seja B uma $n \times n$ matriz sobre F e T o operador linear sobre V que leva X em BX. Determinar o operador T' do Exercício 11.

13. Seja V um espaço vetorial de dimensão finita e f uma forma bilinear simétrica não degenerada sobre V. Associado a f existe um isomorfismo "natural" de V no espaço dual V^*, sendo este isomorfismo a transformação L_f da Seção 10.1. Usando L_f, mostrar que para cada base $\mathscr{B} = (\alpha_1 \ldots, \alpha_n\}$ de V existe uma única base $\mathscr{B}' = \{\alpha_1', \ldots, \alpha_n'\}$ de V tal que $f(\alpha_i, \alpha_j') = \delta_{ij}$. Mostrar então que para todo vetor α em V temos

$$\alpha = \sum_i f(\alpha, \alpha_i')\alpha_i = \sum_i f(\alpha_i, \alpha)\alpha_i'.$$

14. Sejam V, f, $\mathscr{B}$ e $\mathscr{B}'$ como no Exercício 13. Suponhamos que T seja um operador linear sobre V e que T' seja o operador que f associa a T, como no Exercício 11. Mostrar que

(a) $[T']_{\mathscr{B}'} = [T]^t_{\mathscr{B}}$

(b) traço (T) = traço $(T') = \sum_i f(T\alpha_i, \alpha_i')$.

15. Sejam V, f, $\mathscr{B}$ e $\mathscr{B}'$ como no Exercício 13. Suponhamos que $[f]_{\mathscr{B}} = A$. Mostrar que

$$\alpha_i' = \sum_j (A^{-1})_{ij}\alpha_j = \sum_j (A^{-1})_{ji}\alpha_j.$$

16. Seja F um corpo e V o espaço das $n \times 1$ matrizes sobre F. Suponhamos que A seja uma $n \times n$ matriz simétrica inversível sobre F e que f seja a forma bilinear sobre V definida por $f(X, Y) = X^t A Y$. Seja P uma $n \times n$ matriz inversível sobre F e $\mathscr{B}$ a base de V formada pelas colunas de P. Mostrar que a base $\mathscr{B}'$ do Exercício 13 consiste das colunas da matriz $A^{-1}(P^t)^{-1}$.

17. Seja V um espaço vetorial de dimensão finita sobre um corpo F e f uma forma bilinear simétrica sobre V. Para cada subespaço W de V, seja $W^\perp$ o conjunto dos vetores α em V tais que $f(\alpha, \beta) = 0$ para todo β em W. Mostrar que

(a) $W^\perp$ é um espaço.

(b) $V = \{0\}^\perp$.

(c) $V^\perp = \{0\}$ se, e somente se, f é não-degenerada.

(d) posto $(f) = \dim V - \dim V^{\perp}$.

(e) se $\dim V = n$ e $\dim W = m$, então $W^{\perp} \geq n - m$.

(*Sugestão*: Seja $\{\beta_1, \ldots, \beta_n\}$ uma base de W e consideremos a aplicação

$$\alpha \rightarrow (f(\alpha, \beta_1), \ldots, f(\alpha, \beta_m))$$

de V em F^m.)

(f) a restrição de f a W é não-degenerada se, e somente se,

$$W \cap W^{\perp} = \{0\}.$$

(g) $V = W \oplus W^{\perp}$ se, e somente se, a restrição de f a W é não-degenerada.

18. Seja V um espaço vetorial de dimensão finita sobre C e f uma forma bilinear simétrica não-degenerada sobre V. Demonstrar que existe uma base $\mathscr{B}$ de V tal que $\mathscr{B}' = \mathscr{B}$. (Ver o Exercício 13 para uma definição de $\mathscr{B}'$.)

10.3 Formas Bilineares Anti-Simétricas

Em toda esta seção V será um espaço vetorial sobre um subcorpo F do corpo dos números complexos. Uma forma bilinear f sobre V é dita **anti-simétrica** se $f(\alpha, \beta) = -f(\beta, \alpha)$ para todos os vetores α, β em V. Demonstraremos um teorema concernente à simplificação da matriz de uma forma bilinear anti-simétrica sobre um espaço V de dimensão finita. Primeiro faremos algumas observações gerais.

Suponhamos que f seja uma forma bilinear *arbitrária* sobre V. Se fizermos

$$g(\alpha, \beta) = \tfrac{1}{2}[f(\alpha, \beta) + f(\beta, \alpha)]$$
$$h(\alpha, \beta) = \tfrac{1}{2}[f(\alpha, \beta) - f(\beta, \alpha)]$$

então é fácil verificar que g é uma forma bilinear simétrica sobre V e h é uma forma bilinear anti-simétrica sobre V. Além disso $f = g + h$. Ainda mais, esta expressão de f como a soma de uma forma bilinear simétrica e uma anti-simétrica é única. Assim, o espaço $L(V, V, F)$ é a soma direta do subespaço das formas simétricas e o subespaço das formas anti-simétricas.

Se V é de dimensão finita, a forma bilinear f é anti-simétrica se, e somente se, sua matriz A em relação a alguma (ou toda) base ordenada é anti-simétrica, $A^t = -A$. Isto é demonstrado da mesma maneira como se demonstra o fato correspondente sobre formas

bilineares simétricas. Quando f é anti-simétrica, a matriz de f em relação a qualquer base ordenada terá todos os seus elementos diagonais nulos. Isto corresponde exatamente à observação de que $f(\alpha, \alpha) = 0$ para todo α em V, uma vez que $f(\alpha, \alpha) = -f(\alpha, \alpha)$.

Suponhamos que f seja uma forma bilinear anti-simétrica não-nula sobre V. Como $f \neq 0$, existem vetores α, β em V tais que $f(\alpha, \beta) \neq 0$. Multiplicando α por um escalar conveniente, podemos supor que $f(\alpha, \beta) = 1$. Seja γ um vetor arbitrário no subespaço gerado por α e β, digamos, $\gamma = c\alpha + d\beta$. Então

$$\begin{aligned} f(\gamma, \alpha) &= f(c\alpha + d\beta, \alpha) = df(\beta, \alpha) = -d \\ f(\gamma, \beta) &= f(c\alpha + d\beta, \beta) = cf(\alpha, \beta) = c \end{aligned}$$

e, então,

$$\gamma = f(\gamma, \beta)\alpha - f(\gamma, \alpha)\beta. \tag{10-7}$$

Em particular, notemos que α e β são, necessariamente, linearmente independentes; de fato, se $\gamma = 0$, então $f(\gamma, \alpha) = f(\gamma, \beta) = 0$.

Seja W o subespaço bidimensional gerado por α e β. Seja $W^{\perp}$ o conjunto dos vetores δ em V tais que $f(\delta, \alpha) = f(\delta, \beta) = 0$, isto é, o conjunto dos δ tais que $f(\delta, \gamma)$ para todo γ no subespaço W. Afirmamos que $V = W \oplus W^{\perp}$. De fato, seja ε um vetor arbitrário em V e

$$\begin{aligned} \gamma &= f(\varepsilon, \beta)\alpha - f(\varepsilon, \alpha)\beta \\ \delta &= \varepsilon - \gamma. \end{aligned}$$

Então γ está em W e δ está em $W^{\perp}$, pois

$$\begin{aligned} f(\delta, \alpha) &= f(\varepsilon - f(\varepsilon, \beta)\alpha + f(\varepsilon, \alpha)\beta, \alpha) \\ &= f(\varepsilon, \alpha) + f(\varepsilon, \alpha)f(\beta, \alpha) \\ &= 0 \end{aligned}$$

e, analogamente, $f(\delta, \beta) = 0$. Assim, todo ε em V é da forma $\varepsilon = \gamma + \delta$, com γ em W e δ em W^{-}. De (10-7) é evidente que $W \cap W^{\perp} = \{0\}$, portanto, $V = W \oplus W^{\perp}$.

Ora, a restrição de f a $W^{\perp}$ é uma forma bilinear anti-simétrica sobre $W^{\perp}$. Esta restrição pode ser a forma nula. Se não o for, existirão dois vetores α' e β' em $W^{\perp}$ tais que $f(\alpha', \beta') = 1$. Se indicarmos por W' o subespaço bidimensional gerado por α' e β', teremos

$$V = W \oplus W' \oplus W_0$$

onde W_0 é o conjunto dos vetores δ em $W^{\perp}$ tais que $f(\alpha', \delta) = = f(\beta', \delta) = 0$. Se a restrição de f a W_0 não é a forma nula, podemos selecionar vetores α'', β'' em W_0 tais que $f(\alpha'', \beta'') = 1$, e então continuar.

No caso de dimensão finita, deveria estar evidente que obtemos uma seqüência finita de pares de vetores.

$$(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_k, \beta_k)$$

com as seguintes propriedades.

(a) $f(\alpha_j, \beta_j) = 1,\ j = 1, \ldots, k.$

(b) $f(\alpha_i, \alpha_j) = f(\beta_i, \beta_j) = f(\alpha_i, \beta_j) = 0,\ i \neq j.$

(c) Se W_j é o subespaço bidimensional gerado por α_j e β_j, então

$$V = W_1 \oplus \ldots \oplus W_k \oplus W_0$$

onde todo vetor em W_0 é 'ortogonal' a todos α_j e β_j, e a restrição de f a W_0 é a forma nula.

Teorema 6. *Seja* V *um espaço vetorial* n-*dimensional sobre um subcorpo do corpo dos números complexos e seja* f *uma forma bilinear anti-simétrica sobre* V. *Então o posto* r *de* f *é par e se* r = 2k, *existe uma base ordenada de* V *em relação à qual a matriz de* f *é a soma direta da* (n − r) × (n − r) *matriz nula e* k *cópias da* 2 × 2 *matriz*

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Demonstração. Sejam $\alpha_1, \beta_1, \ldots, \alpha_k, \beta_k$ vetores que satisfaçam as condições (a), (b) e (c) acima. Seja $\{\gamma_1, \ldots, \gamma_s\}$ uma base ordenada arbitrária do subespaço W_0. Então

$$\mathscr{B} = \{\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_k, \beta_k, \gamma_1, \ldots, \gamma_s\}$$

é uma base ordenada de V. De (a), (b) e (c) é evidente que a matriz de f em relação à base ordenada $\mathscr{B}$ é a soma direta da $(n - 2k) \times \times (n - 2k)$ matriz nula e k cópias da 2×2 matriz

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{10-8}$$

Além disso, é evidente que o posto desta matriz, e, portanto, o posto de f, é $2k$.

Uma conseqüência disto acima é que se f é uma forma bilinear anti-simétrica não-degenerada sobre V, então a dimensão de V deve ser par. Se $\dim V = 2k$, existe uma base ordenada $\{\alpha_1, \beta_1, \ldots, \alpha_k, \beta_k\}$ de V tal que

$$f(\alpha_i, \beta_j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

$$f(\alpha_i, \alpha_j) = f(\beta_i, \beta_j) = 0.$$

A matriz de f em relação a esta base ordenada é a soma direta de k cópias da 2×2 matriz anti-simétrica (10-8). Obtém-se uma outra forma canônica para a matriz de uma forma anti-simétrica não-degenerada se, ao invés da base ordenada acima, considera-se a base ordenada

$$\{\alpha_1, \ldots, \alpha_k, \beta_k, \ldots, \beta_1\}.$$

O leitor deverá achar fácil verificar que a matriz de f em relação à última base ordenada é da forma em blocos

$$\begin{bmatrix} 0 & J \\ -J & 0 \end{bmatrix}$$

onde J é a $k \times k$ matriz

$$\begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ 1 & \cdots & 0 & 0 \end{bmatrix}.$$

Exercícios

1. Seja V um espaço vetorial sobre um corpo F. Mostrar que o conjunto das formas bilineares anti-simétricas sobre V é um subespaço de $L(V, V, F)$.

2. Determinar todas as formas bilineares anti-simétricas sobre R^3.

3. Determinar uma base do espaço das formas bilineares anti simétricas sobre R^n.

4. Seja f uma forma bilinear simétrica sobre C^n e g uma forma bilinear anti-simétrica sobre C^n. Suponhamos que $f+g=0$. Mostrar que $f=g=0$.

5. Seja V um espaço vetorial n-dimensional sobre um subcorpo F de C. Demonstrar o seguinte:

(a) A equação $(Pf)(\alpha, \beta) = \frac{1}{2}f(\alpha, \beta) - \frac{1}{2}f(\beta, \alpha)$ define um operador linear P sobre $L(V, V, F)$.

(b) $P^2 = P$, isto é, P é uma projeção.

(c) posto $(P) = \dfrac{n(n-1)}{2}$; nulidade $(P) = \dfrac{n(n+1)}{2}$.

(d) Se U é um operador linear sobre V, a equação $(U^+f)(\alpha, \beta) = f(U\alpha, U\beta)$ define um operador U^+ sobre $L(V, V, F)$.

(e) Para todo operador linear U, a projeção P comuta com U^+.

6. Demonstrar um análogo do Exercício 11 na Seção 10.2 para formas bilineares anti-simétricas não-degeneradas.

7. Seja f uma forma bilinear sobre um espaço vetorial V. Sejam L_f e R_f as aplicações de V em V^* associadas a f na Seção 10.1. Demonstrar que f é anti-simétrica se, e somente se, $L_f = -R_f$.

8. Demonstrar um análogo do Exercício 17 na Seção 10.2 para formas anti-simétricas.

9. Seja V um espaço vetorial de dimensão finita e L_1, L_2 funcionais lineares sobre V. Mostrar que a equação

$$f(\alpha, \beta) = L_1(\alpha)L_2(\beta) - L_1(\beta)L_2(\alpha)$$

define uma forma bilinear anti-simétrica sobre V. Mostrar que $f=0$ se, e somente se, L_1 e L_2 são linearmente dependentes.

10. Seja V um espaço vetorial de dimensão finita sobre um subcorpo do corpo dos números complexos e f uma forma bilinear anti-simétrica sobre V. Mostrar que f tem posto 2 se, e somente se, existem funcionais lineares linearmente independentes L_1, L_2, sobre V tais que

$$f(\alpha, \beta) = L_1(\alpha)L_2(\beta) - L_1(\beta)L_2(\alpha).$$

11. Seja f uma forma bilinear anti-simétrica arbitrária sobre R^3. Demonstrar que existem funcionais lineares L_1, L_2 tais que

$$f(\alpha, \beta) = L_1(\alpha) L_2(\beta) - L_1(\beta) L_2(\alpha).$$

12. Seja V um espaço vetorial de dimensão finita sobre um subcorpo do corpo dos números complexos e sejam f, g formas bilineares anti-simétricas sobre V. Mostrar que existe um operador linear *inversível* T sobre V tal que $f(T\alpha, T\beta) = g(\alpha, \beta)$ para todos α, β se, e somente se, f e g têm o mesmo posto.

13. Mostrar que o resultado do Exercício 12 é válido para formas bilineares simétricas sobre um espaço vetorial complexo, mas não é válido para formas bilineares simétricas sobre um espaço vetorial real.

10.4 Grupos que Conservam Formas Bilineares

Seja f uma forma bilinear sobre o espaço vetorial V e seja T um operador linear sobre V. Dizemos que T **conserva** f se $f(T\alpha, T\beta) = f(\alpha, \beta)$ para todos α, β em V. Para quaisquer T e f, a função g definida por $g(\alpha, \beta) = f(T\alpha, T\beta)$, como se vê facilmente, é uma forma bilinear sobre V. Dizer que T conserva f é simplesmente dizer que $g = f$. O operador idêntico conserva toda forma bilinear. Se S e T são operadores lineares que conservam f, o produto ST também conserva f; de fato, $f(ST\alpha, ST\beta) = f(T\alpha, T\beta) = f(\alpha, \beta)$.

Em outras palavras, a coleção de operadores lineares que conservam uma dada forma bilinear é fechada em relação à formação de produtos (de operadores). Em geral, não se pode dizer muito mais acerca desta coleção de operadores; no entanto, se f é não-degenerada, temos o que segue.

Teorema 7. *Seja* f *uma forma bilinear não-degenerada sobre um espaço vetorial* V *de dimensão finita. O conjunto dos operadores lineares sobre* V *que conservam* f *é um grupo em relação à operação de composição.*

Demonstração. Seja G o conjunto dos operadores lineares que conservam f. Observamos que o operador idêntico está em G e que, sempre que S e T estão em G, o composto ST também está em G. A partir do fato de que f é não-degenerada, demonstraremos que todo operador T em G é inversível e que T^{-1} também está em G. Suponhamos que T conserve f. Seja α um vetor no núcleo de T. Então, para todo β em V, temos

$$f(\alpha, \beta) = f(T\alpha, T\beta) = f(0, T\beta) = 0.$$

Como f é não-degenerada, $\alpha = 0$. Assim, T é inversível. Evidentemente T^{-1} também conserva f, pois

$$f(T^{-1}\alpha, T^{-1}\beta) = f(TT^{-1}\alpha, TT^{-1}\beta) = f(\alpha, \beta).$$

Se f é uma forma bilinear não-degenerada sobre o espaço V de dimensão finita, então cada base ordenada $\mathscr{B}$ de V determina um grupo de matrizes que "conservam" f. O conjunto de todas as matrizes $[T]_{\mathscr{B}}$, onde T é um operador linear que conserva f, será um grupo em relação à multiplicação de matrizes. Existe uma descrição alternativa deste grupo de matrizes, como segue. Seja $A = [f]_{\mathscr{B}}$, de modo que se α e β são vetores em V com respectivas matrizes de coordenadas X e Y em relação a $\mathscr{B}$, teremos

$$f(\alpha, \beta) = X^t AY.$$

Seja T um operador linear arbitrário sobre V e $M = [T]_{\mathscr{B}}$. Então

$$\begin{aligned} f(T\alpha, T\beta) &= (MX)^t A(MY) \\ &= X^t(M^t AM)Y. \end{aligned}$$

Conseqüentemente, T conserva f se, e somente se, $M^t AM = A$. Em linguagem matricial, o Teorema 7 diz o seguinte: Se A é uma $n \times n$ matriz inversível, o conjunto das $n \times n$ matrizes M tais que $M^t AM = A$ é um grupo em relação à multiplicação matricial. Se $A = [f]_{\mathscr{B}}$, então M está neste grupo de matrizes se, e somente se, $M = [T]_{\mathscr{B}}$ onde T é um operador linear que conserva f.

Antes de passarmos a alguns exemplos, façamos mais uma observação. Suponhamos que f seja uma forma bilinear que seja simétrica. Um operador linear T conserva f se, e somente se, T conserva a forma quadrática

$$q(\alpha) = f(\alpha, \alpha)$$

associada a f. Se T conserva f, certamente temos

$$q(T\alpha) = f(T\alpha, T\alpha) = f(\alpha, \alpha) = q(\alpha)$$

para todo α em V. Reciprocamente, como f é simétrica, a identidade de polarização

$$f(\alpha, \beta) = \tfrac{1}{4}q(\alpha + \beta) - \tfrac{1}{4}q(\alpha - \beta)$$

nos mostra que T conserva f se $q(T\gamma) = q(\gamma)$ para cada γ em V. (Estamos supondo aqui que o corpo de escalares seja um subcorpo do corpo dos números complexos.)

Exemplo 6. Seja V o espaço R^n ou o espaço C^n. Seja f a forma bilinear

$$f(\alpha, \beta) = \sum_{j=1}^{n} x_j y_j$$

onde $\alpha = (x_1, \ldots, x_n)$ e $\beta = (y_1, \ldots, y_n)$. O grupo que conserva f é denominado o **grupo ortogonal** (real ou complexo) n-dimensional. O nome 'grupo ortogonal' é mais comumente aplicado ao grupo associado de matrizes em relação à base ordenada canônica. Como a matriz de f em relação à base canônica é I, este grupo consiste das matrizes M que satisfazem $M^tM = I$. Uma tal matriz M é dita uma $n \times n$ **matriz ortogonal** (real ou complexa). Os dois $n \times n$ grupos ortogonais são usualmente indicados por $O(n, R)$ e $O(n, C)$. Evidentemente, o grupo ortogonal é também o grupo que conserva a forma quadrática

$$q(x_1, \ldots, x_n) = x_1^2 + \ldots + x_n^2.$$

Exemplo 7. Seja f a forma bilinear simétrica sobre R^n com forma quadrática

$$q(x_1, \ldots, x_n) = \sum_{j=1}^{p} x_j^2 - \sum_{j=p+1}^{n} x_j^2.$$

Então f é não-degenerada e tem assinatura $2p - n$. O grupo das matrizes que conservam uma forma deste tipo é denominado um **grupo pseudo-ortogonal.** Quando $p = n$, obtemos o grupo ortogonal $O(n, R)$ como um tipo particular de grupo pseudo-ortogonal. Para cada um dos $n + 1$ valores $p = 0, 1, 2, \ldots, n$, obtemos uma forma bilinear diferente f; contudo, para $p = k$ e $p = n - k$ as formas são uma a oposta da outra e têm, portanto, o mesmo grupo associado. Assim, quando n é ímpar, temos $(n + 1)/2$ grupos pseudo-ortogonais de $n \times n$ matrizes e quando n é par, temos $(n + 2)/2$ desses grupos.

Teorema 8. *Seja* V *um espaço vetorial* n*-dimensional sobre o corpo dos números complexos e seja* f *uma forma bilinear simétrica*

não-degenerada sobre V. *Então o grupo que conserva* f *é isomorfo ao grupo ortogonal complexo* O(n, C).

Demonstração. Obviamente, por um isomorfismo entre grupos, queremos dizer uma correspondência bijetora entre seus elementos que 'conserva' a operação de grupo. Seja G o grupo dos operadores lineares sobre V que conservam a forma bilinear f. Como f é simétrica e não-degenerada, o Teorema 4 nos diz que existe uma base ordenada $\mathscr{B}$ de V em relação à qual f é representada pela $n \times n$ matriz unidade. Portanto, um operador linear T conserva f se, e somente se, sua matriz em relação à base ordenada $\mathscr{B}$ é uma matriz ortogonal complexa. Logo,

$$T \to [T]_{\mathscr{B}}$$

é um isomorfismo de G em $O(n, C)$.

Teorema 9. *Seja* V *um espaço vetorial* n-*dimensional sobre o corpo dos números reais e seja* f *uma forma bilinear simétrica não-degenerada sobre* V. *Então, o grupo que conserva* f *é isomorfo a um* n × n *grupo pseudo-ortogonal.*

Demonstração. Repetir a demonstração do Teorema 8, usando o Teorema 5 em vez do Teorema 4.

Exemplo 8. Seja f a forma bilinear simétrica sobre R^4 com forma quadrática

$$q(x, y, z, t) = t^2 - x^2 - y^2 - z^2.$$

Um operador linear T sobre R^4 que conserva esta forma bilinear (ou quadrática) particular é denominado uma **transformação de Lorentz** e o grupo que conserva f é dito o **grupo de Lorentz**. Gostaríamos de dar um método para a descrição de algumas transformações de Lorentz.

Seja H o espaço vetorial real das 2×2 matrizes complexas A que sejam hermitianas, $A = A^*$. É fácil verificar que

$$\Phi(x, y, z, t) = \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

define um isomorfismo Φ de R^4 no espaço H. Por meio deste isomorfismo, a forma quadrática q é levada sobre a função determinante, isto é,

$$q(x, y, z, t) = \det \begin{bmatrix} t + x & y + iz \\ y - iz & t - x \end{bmatrix}$$

ou

$$q(\alpha) = \det \Phi(\alpha).$$

Isto sugere que podemos estudar as transformações de Lorentz sobre R^4 estudando operadores lineares sobre H que conservam determinantes.

Seja M uma 2×2 matriz complexa arbitrária e, para uma matriz hermitiana A, definamos

$$U_M(A) = MAM^*.$$

Ora, MAM^* também é hermitiana. A partir disto, é fácil ver que U_M é um operador linear (real) sobre H. Perguntamos quando é que U_M 'conserva' determinantes, isto é, $\det [U_M(A)] = \det A$ para cada A em H. Como o determinante de M^* é o complexo conjugado do determinante de M, vemos que

$$\det [U_M(A)] = |\det M|^2 \det A.$$

Assim, U_M conserva determinantes exatamente quando $\det M$ tem valor absoluto 1.

Selecionemos, então, uma 2×2 matriz complexa arbitrária M para a qual $|\det M| = 1$. Então, U_M é um operador linear sobre H que conserva determinantes. Definamos

$$T_M = \Phi^{-1} U_M \Phi.$$

Como Φ é um isomorfismo, T_M é um operador linear sobre R^4. Além disso, T_M é uma transformação de Lorentz, pois

$$\begin{aligned} q(T_M\alpha) &= q(\Phi^{-1} U_M \Phi\alpha) \\ &= \det (\Phi\Phi^{-1} U_M \Phi\alpha) \\ &= \det (U_M \Phi\alpha) \\ &= \det (\Phi\alpha) \\ &= q(\alpha) \end{aligned}$$

e, portanto, T_M conserva a forma quadrática q.

Usando 2×2 matrizes particulares M, pode-se usar o método acima para calcular transformações de Lorentz particulares. Dois

comentários devem ser feitos neste ponto; eles não são difíceis de serem verificados.

(1) Se M_1 e M_2 são 2×2 matrizes inversíveis com elementos complexos, então $U_{M_1} = U_{M_2}$ se, e somente se, M_2 é um múltiplo escalar de M_1. Assim, todas as transformações de Lorentz acima exibidas podem ser obtidas a partir de matrizes unimodulares M, isto é, a partir de matrizes M que satisfazem $\det M = 1$. Se M_1 e M_2 são matrizes unimodulares tais que $M_1 \neq M_2$ e $M_1 \neq -M_2$, então, $T_{M_1} \neq T_{M_2}$.

(2) Nem toda transformação de Lorentz pode ser obtida pelo método acima.

Exercícios

1. Seja M um elemento do grupo ortogonal complexo, $O(n, C)$. Mostrar que M^t, $\overline{M}$ e $M^* = \overline{M}^t$ também pertencem a $O(n, C)$.

2. Suponhamos que M pertença a $O(n, C)$ e que M' seja semelhante a M. M' também pertence a $O(n, C)$?

3. Seja

$$y_j = \sum_{k=1}^{n} M_{jk} x_k$$

onde M é um elemento de $O(n, C)$. Mostrar que

$$\sum_j y_j^2 = \sum_j x_j^2.$$

4. Seja M uma $n \times n$ matriz sobre C com colunas $M_1, M_2, \ldots, M_n$. Mostrar que M pertence a $O(n, C)$ se, e somente se,

$$M_j^t M_k = \delta_{jk}.$$

5. Seja X uma $n \times 1$ matriz sobre C. Em que condições $O(n, C)$ contém uma matriz M cuja primeira coluna seja X?

6. Determinar uma matriz em $O(3, C)$ cuja primeira linha seja $(2i, 2i, 3)$.

7. Seja V o espaço das $n \times 1$ matrizes sobre C e f a forma bilinear sobre V dada por $f(X, Y) = X^t Y$. Seja M pertencente a $O(n, C)$. Qual é a matriz de f em relação à base de V formada pelas colunas $M_1, M_2, \ldots, M_n$ de M?

8. Seja X uma $n \times 1$ matriz sobre C tal que $X^t X = 1$ e seja I_j a j-ésima coluna da matriz unidade. Mostrar que existe uma matriz M em $O(n, C)$ tal que $MX = I_j$. Se X tem elementos reais mostrar que existe uma M em $O(n, R)$ com a propriedade de que $MX = I_j$.

9. Seja V o espaço das $n \times 1$ matrizes sobre C, A uma $n \times n$ matriz sobre C e f a forma bilinear sobre V dada por $f(X, Y) = X^t A Y$. Mostrar que f é invariante sob $O(n, C)$, isto é, $f(MX, MY) = f(X, Y)$ para todas X, Y em V e toda M em $O(n, C)$, se, e somente se, A comuta com cada elemento de $O(n, C)$.

10. Seja S um conjunto arbitrário de $n \times n$ matrizes sobre C e S' o conjunto das $n \times n$ matrizes sobre C que comutam com todo elemento de S. Mostrar que S' é uma álgebra sobre C.

11. Seja F um subcorpo de C, V um espaço vetorial de dimensão finita sobre F e f uma forma bilinear não-singular sobre V. Se T é um operador linear sobre V que conserva f, demonstrar que $\det T = \pm 1$.

12. Seja F um subcorpo de C, V o espaço das $n \times 1$ matrizes sobre F, A uma $n \times n$ matriz inversível sobre F e f a forma bilinear sobre V dada por $f(X, Y) = X^t A Y$. Se M é uma $n \times n$ matriz sobre F, mostrar que M conserva f se, e somente se, $A^{-1} M^t A = M^{-1}$.

13. Seja g uma forma bilinear não-singular sobre um espaço vetorial V de dimensão finita. Suponhamos que T seja um operador linear inversível sobre V e que f seja a forma bilinear sobre V dada por $f(\alpha, \beta) = g(\alpha, T\beta)$. Se U é um operador linear sobre V, determinar condições necessárias e suficientes para que U conserve f.

14. Seja T um operador linear sobre C^2 que conserva a forma quadrática $x_1^2 - x_2^2$. Mostrar que

(a) $\det(T) = \pm 1$.

(b) Se M é a matriz de T em relação à base canônica, então $M_{22} = \pm M_{11}$, $M_{21} = \pm M_{12}$, $M_{11}^2 - M_{12}^2 = 1$.

(c) Se $\det M = 1$, então existe um número complexo não-nulo c tal que

$$M = \frac{1}{2}\begin{bmatrix} c + \frac{1}{c} & c - \frac{1}{c} \\ c - \frac{1}{c} & c + \frac{1}{c} \end{bmatrix}.$$

(d) Se $\det M = -1$, então existe um número complexo c tal que

$$M = \frac{1}{2}\begin{bmatrix} c + \frac{1}{c} & - c - \frac{1}{c} \\ -c + \frac{1}{c} & -c - \frac{1}{c} \end{bmatrix}.$$

15. Seja f a forma bilinear sobre C^2 definida por

$$f((x_1, x_2), (y_1, y_2)) = x_1 y_2 - x_2 y_1.$$

Mostrar que

(a) Se T é um operador linear sobre C^2, então $f(T\alpha, T\beta) = (\det T) f(\alpha, \beta)$ para todos α, β em C^2.

(b) T conserva f se, e somente se, $\det T = +1$.

(c) O que é que (b) diz acerca do grupo das 2×2 matrizes M tais que $M^t A M = A$ onde

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}?$$

16. Seja n um inteiro positivo, I a $n \times n$ matriz unidade sobre C e J a $2n \times 2n$ matriz dada por

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}.$$

Seja M a $2n \times 2n$ matriz sobre C da forma

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

onde A, B, C, D são $n \times n$ matrizes sobre C. Determinar condições necessárias e suficientes sobre A, B, C, D para que $M^t J M = J$.

17. Determinar todas as formas bilineares sobre o espaço das $n \times 1$ matrizes sobre R que sejam invariantes sob $O(n, R)$.

18. Determinar todas as formas bilineares sobre o espaço das $n \times 1$ matrizes sobre C que sejam invariantes sob $O(n, C)$.

APÊNDICE

Este Apêndice divide-se de maneira lógica em duas partes. A primeira parte, compreendendo as três primeira seções, contém certos conceitos fundamentais que ocorrem por todo o livro (na verdade, por toda a matemática). É mais uma espécie de introdução do livro que apêndice. A segunda parte é mais genuinamente um apêndice ao texto.

A Seção 1 contém uma discussão sobre conjuntos, suas reuniões e interseções. A Seção 2 discute o conceito de função e as idéias afins de imagem, domínio, função inversa e a restrição de uma função a um subconjunto do seu domínio. A Seção 3 trata das relações de equivalência. O assunto destas três seções, especialmente o das Seções 1 e 2, é apresentado de uma maneira bem concisa. É tratado mais como um acordo sobre a terminologia que como uma exposição detalhada. Num sentido lógico estrito, esta matéria constitui uma parte dos pré-requisitos para a leitura do livro; contudo, o leitor não deverá se desencorajar se não conseguir aprender completamente o significado das idéias na sua primeira leitura. Estas idéias são importantes, mas o leitor que não tiver muita familiaridade com elas deverá achar mais fácil absorvê-las se rever a discussão de tempos em tempos, à medida que for lendo o texto em si.

As Seções 4 e 5 consideram as relações de equivalência no contexto da álgebra linear. A Seção 4 contém uma discussão breve de espaços quocientes. Pode ser lida a qualquer momento após os dois ou três primeiros capítulos do livro. A Seção 5 considera rapidamente algumas das relações de equivalência que aparecem no livro, tentando indicar como alguns dos resultados do livro poderiam ser interpretados do ponto de vista de relações de equivalência. A Seção 6 descreve o Axioma da Escolha e suas implicações para a Álgebra Linear.

A.1 Conjuntos

Usaremos as palavras 'conjunto', 'classe', 'coleção' e 'família' indiferentemente, apesar de darmos preferência a 'conjunto'. Se S é um conjunto e x é um objeto do conjunto S, diremos que x é um **elemento de** S, que x **pertence** a S ou simplesmente que x está em S. Se S possui apenas um número finito de elementos, $x_1, \ldots, x_n$, freqüentemente descrevemos S exibindo seus elementos dentro de chaves:

$$S = \{x_1, \ldots, x_n\}.$$

Assim, o conjunto S dos inteiros positivos de 1 a 5, seria

$$S = \{1, 2, 3, 4, 5\}.$$

Se S e T são conjuntos, dizemos que S é um **subconjunto de** T, ou que S está **contido em** T, se cada elemento de S é um elemento de T. Cada conjunto S é um subconjunto de si mesmo. Se S é um subconjunto de T mas S e T não são idênticos, denominamos S um **subconjunto próprio** de T. Em outras palavras, S é um subconjunto próprio de T se S está contido em T mas T não está contido em S.

Se S e T são conjuntos, a **reunião de** S **com** T é o conjunto $S \cup T$, constituído de todos os objetos x que são elementos de S ou de T. **A interseção de** S **com** T é o conjunto $S \cap T$, formado por todos os x que são elementos de S e de T. Para dois conjuntos arbitrários, S e T, a interseção $S \cap T$ é um subconjunto da reunião $S \cup T$. Isto deve auxiliar a esclarecer o uso da palavra 'ou' que prevalecerá neste livro: Quando dizemos que x está em S ou em T, não excluímos a possibilidade de x estar em ambos S e T.

Para que a interseção de S e T seja sempre um conjunto, é necessário introduzir o **conjunto vazio**, isto é, o conjunto sem elementos. Então $S \cap T$ é o conjunto vazio se, e somente se, S e T não têm elementos em comum.

Freqüentemente precisaremos discutir a reunião ou a interseção de diversos conjuntos. Se $S_1, \ldots, S_n$ são conjuntos, sua **reunião** é o conjunto $\bigcup_{j=1}^{n} S_j$ formado por todos os x que são elementos de pelo menos um dos conjuntos $S_1, \ldots, S_n$. Sua **interseção** é o conjunto $\bigcap_{j=1}^{n} S_j$, formado por todos os x que são elementos

de cada um dos conjuntos $S_1, \ldots, S_n$. Em algumas ocasiões, discutiremos a reunião ou a interseção de uma coleção infinita de conjuntos. Deveria estar evidente a maneira como tais reuniões e interseções são definidas. O exemplo que segue deverá esclarecer estas definições e uma notação para elas.

Exemplo 1. Indiquemos por R o conjunto dos (de todos os) números reais (a reta real). Se t está em R, associamos a t um subconjunto S_t de R, definido como segue: S_t consiste dos números reais x que não são menores que t.

(a) $S_{t_1} \cup S_{t_2} = S_t$ onde t é o menor entre t_1 e t_2.

(b) $S_{t_1} \cap S_{t_2} = S_t$, onde t é o maior entre t_1 e t_2.

(c) Seja I o intervalo unitário, isto é, o conjunto dos t em R que satisfazem $0 \leq t \leq 1$. Então

$$\bigcup_{t \text{ em } I} S_t = S_0$$

$$\bigcap_{t \text{ em } I} S_t = S_1$$

A.2 Funções

Uma **função** consiste do seguinte:

(1) um conjunto X, denominado o domínio da função;

(2) um conjunto Y, denominado o contradomínio da função;

(3) uma regra (ou correspondência) f, que associa a cada elemento x de X um único elemento $f(x)$ de Y.

Se (X, Y, f) é uma função, diremos que **f é uma função de X em Y**. Isto é um tanto confuso, pois não é f que é a função; f é a regra da função. No entanto, este uso do mesmo símbolo para a função e sua regra nos fornece uma maneira muito mais maleável de falar sobre funções. Assim, diremos que f é uma função de X em Y, que X é o domínio de f e que Y é o contradomínio de f — tudo isto significando que (X, Y, f) é uma função como definido acima. Existem várias outras palavras que são comumente usadas no lugar da palavra 'função'. Algumas delas são 'transformação' 'operador' e 'aplicação'. Estas são usadas em contextos onde pareçam ser mais sugestivas na transmissão do papel desempenhado por uma função específica.

Se f é uma função de X em Y, a **imagem** de f é o conjunto dos $f(x)$, x em X. Em outras palavras, a imagem de f consiste dos elementos y em Y tais que $y = f(x)$, para algum x em X. Se a imagem de f é todo o conjunto Y, dizemos que f é uma **função sobrejetora de** X **em** Y ou simplesmente que f é **sobrejetora.** A imagem de f é freqüentemente indicada por $f(X)$.

Exemplo 2. (a) Seja X o conjunto dos números reais e seja $Y = X$. Seja f a função de X em Y definida por $f(x) = x^2$. A imagem de f é o conjunto dos números reais não-negativos. Assim, f não é sobrejetora.

(b) Seja X o plano euclidiano e $Y = X$. Seja f definida como segue: se P é um ponto do plano, então $f(P)$ é o ponto obtido girando-se P de 90° (em torno da origem, no sentido anti-horário). A imagem de f é todo Y, isto é, todo o plano; portanto, f é sobrejetora.

(c) Novamente, seja X o plano euclidiano. Coloquemos coordenadas em X como em geometria analítica, usando duas retas perpendiculares para identificarmos os pontos de X com pares ordenados de números reais (x_1, x_2). Seja Y o eixo dos x_1, isto é, o conjunto dos pontos (x_1, x_2) com $x_2 = 0$. Se P é um ponto de X, seja $f(P)$ ponto obtido projetando P sobre o eixo dos x_1, paralelamente ao eixo dos x_2. Em outras palavras, $f((x_1, x_2)) = (x_1, 0)$. A imagem de f é todo Y, portanto f é sobrejetora.

(d) Seja X o conjunto dos números reais e seja Y o conjunto dos números reais positivos. Definamos uma função f de X em Y por $f(x) = e^x$. Então, f é uma função sobrejetora de X em Y.

(e) Seja X o conjunto dos números reais positivos e Y o conjunto dos números reais. Seja f a função logarítmica natural, isto é, a função definida por $f(x) = \log x = \ln x$. Novamente, f é sobrejetora, isto é, todo número real é o logaritmo natural de algum número positivo.

Suponhamos que X, Y e Z sejam conjuntos, que f seja uma função de X em Y e que g seja uma função de Y em Z. Existe, associada a f e g, uma função $g \bigcirc f$ de X em Z, conhecida como a **composta** de g e f. É definida por

$$(g \bigcirc f)(x) = g(f(x)).$$

Para um exemplo simples, seja $X = Y = Z$, o conjunto dos números reais, sejam f, g, h as funções de X em X definidas por

$$f(x) = x^2, \; g(x) = e^x, \; h(x) = e^{x^2}$$

e, então, $h = g \bigcirc f$. A composta $g \bigcirc f$ é freqüentemente indicada por gf; contudo, como mostra o exemplo simples acima, existem ocasiões em que isto pode levar a confusão.

Uma questão de interesse é a que segue. Suponhamos que f seja uma função de X em Y. Quando é que existe uma função g de Y em X tal que $g(f(x)) = x$ para todo x em X? Indicando por I a **função idêntica** sobre X, isto é, a função de X em X definida por $I(x) = x$, estamos perguntando: Quando é que existe uma função g de Y em X tal que $g \bigcirc f = I$? A grosso modo, queremos uma função g que 'leva cada elemento de Y de volta ao lugar de onde veio'. Para que uma tal g exista, f precisa ser **injetora**, isto é, f precisa ter a propriedade de que se $x_1 \neq x_2$, então $f(x_1) \neq f(x_2)$. Se f é injetora, existe uma tal função g. Ela é definida como segue: Seja y um elemento de Y. Se y está na imagem de f, então existe um elemento x em X tal que $y = f(x)$; como f é injetora, existe exatamente um tal x. Definamos $g(y) = x$. Se y não está na imagem de f, definamos $g(y)$ como sendo um elemento qualquer de X. Obviamente, temos $g \bigcirc f = I$.

Seja f uma função de X em Y. Dizemos que f é **inversível** se existe uma função g de Y em X tal que

(1) $g \bigcirc f$ é a função idêntica sobre X,

(2) $f \bigcirc g$ é a função idêntica sobre Y.

Acabamos de ver que se existe uma g satisfazendo (1), então f é injetora. Analogamente, pode-se ver que se existe uma g satisfazendo (2), a imagem de f é todo o conjunto Y, isto é, f é sobrejetora. Assim, se f é inversível, f é injetora e sobrejetora (bijetora). Reciprocamente, se f é bijetora, existe uma função g de Y em X que satisfaz (1) e (2). Além disso, esta g é única. Ela é a função de Y em X, definida por esta regra: se y está em Y, então $g(y)$ é o único elemento x em X para o qual $f(x) = y$.

Se f é inversível (bijetora) a **inversa** de f é a única função f^{-1} de Y em X que satisfaz

(1′) $f^{-1}(f(x)) = x$, para cada x em X,

(2′) $f(f^{-1}(y)) = y$, para cada y em Y.

Exemplo 3. Consideremos as funções do Exemplo 2.

(a) Se $X = Y$, o conjunto dos números reais, e $f(x) = x^2$, então f não é inversível. De fato, f não é injetora nem sobrejetora.

(b) Se $X = Y$, o espaço euclidiano, e f é a 'rotação de 90°', então f é bijetora. A função inversa f^{-1} é a 'rotação de $-90°$' ou a 'rotação de 270°'.

(c) Se X é o plano, Y o eixo dos x_1 e $f((x_1, x_2)) = (x_1, 0)$, então f não é inversível. De fato, apesar de ser sobrejetora, f não é injetora.

(d) Se X é o conjunto dos números reais, Y o conjunto dos números reais positivos e $f(x) = e^x$, então f é inversível. A função f^{-1} é a função logarítmica natural da parte (e): $\log e^x = x$, $e^{\log y} = y$.

(e) A inversa desta função logarítmica natural é a função exponencial da parte (d).

Seja f uma função de X em Y e seja f_0 uma função de X_0 em Y_0. Dizemos que f_0 é uma **restrição** de f (ou uma restrição de f a X_0) se

(1) X_0 é um subconjunto de X,

(2) $f_0(x) = f(x)$, para cada x em x_0.

Evidentemente, quando f_0 é uma restrição de f, decorre que Y_0 é um subconjunto de Y. O nome 'restrição' vem do fato de que f e f_0 têm a mesma regra e diferem principalmente porque restringimos o domínio de definição da regra ao subconjunto X_0 de X.

Se nos é dada uma função f e um subconjunto arbitrário X_0 de X, existe uma maneira óbvia de construir uma restrição de f a X_0. Definamos uma função f_0 de X_0 em Y por $f_0(x) = f(x)$ para cada x em X_0. Poder-se-ia perguntar por que não denominamos esta f_0 a restrição de f a X_0. A razão é que, ao discutirmos restrições de f, queremos a liberdade de mudar o contradomínio Y, bem como o domínio de X.

Exemplo 4. (a) Seja X o conjunto dos números reais e f a função de X em X definida por $f(x) = x^2$. Então, f não é uma função inversível mas o será se restringirmos seu domínio aos números reais não-negativos. Seja X_0 o conjunto dos números reais não-negativos e seja f_0 a função de X_0 em X_0 definida por $f_0(x) = x^2$. Então f_0 é uma restrição de f a X_0. Ora, f não é injetora nem sobrejetora, enquanto que f_0 é injetora e sobrejetora. A última afirmação diz simplesmente que cada número não-negativo é o qua-

drado de exatamente um número não-negativo. A função inversa f_0^{-1} é a função de X_0 em X_0 definida por $f_0^{-1}(x) = \sqrt{x}$.

(b) Seja X o conjunto dos números reais e seja f a função de X em X definida por $f(x) = x^3 + x^2 + 1$. A imagem de f é todo X, portanto, f é sobrejetora. A função f certamente não é injetora, pois, $f(-1) = f(0)$. Mas f é injetora sobre X_0, o conjunto dos números reais não-negativos, pois a derivada de f é positiva para $x > 0$. Quando x percorre todos os números reais não-negativos, $f(x)$ percorre todos os números reais y tais que $y \geq 1$. Se indicarmos por Y_0 o conjunto dos $y \geq 1$ e por f_0 a função de X_0 em Y_0 definida por $f_0(x) = f(x)$, então f_0 é uma função bijetora de X_0 em Y_0. Conseqüentemente, f_0 possui uma função inversa f_0^{-1} de Y_0 em X_0. Qualquer fórmula para $f_0^{-1}(y)$ é bastante complicada.

(c) Novamente, seja X o conjunto dos números reais e seja f a função seno, isto é, a função de X em X definida por $f(x) = \text{sen } x$. A imagem de f é o conjunto dos y tais que $-1 \leq y \leq 1$; logo, f não é sobrejetora. Como $f(x + 2\pi) = f(x)$, vemos que f não é injetora. Se indicarmos por X_0 o intervalo $-\pi/2 \leq x \leq \pi/2$, então f é injetora em X_0. Seja Y_0 o intervalo $-1 \leq y \leq 1$ e seja f_0 a função de X_0 em Y_0 definida por $f_0(x) = \text{sen } x$. Então f_0 é uma restrição de f ao intervalo X_0 e f_0 é bijetora. Esta é apenas uma outra maneira de dizer que, no intervalo de $-\pi/2$ a $\pi/2$, a função seno toma cada valor entre -1 e 1 exatamente uma vez. A função f_0^{-1} é a função arco-seno:

$$f_0^{-1}(y) = \text{sen}^{-1} y = \text{arc sen } y.$$

(d) Este é um exemplo geral de uma restrição de uma função. É muito mais característico do tipo de restrição que usaremos neste livro do que os exemplos em (b) e (c) acima. O exemplo em (a) é um caso particular deste. Seja X um conjunto e f uma função de X em si mesmo. Seja X_0 um subconjunto de X. Dizemos que X_0 **é invariante sob** f se para cada x em X_0 o elemento $f(x)$ está em X_0. Se X_0 é invariante sob f, então f induz uma função f_0 de X_0 em si mesmo, restringindo-se o seu domínio de definição a X_0. A importância da invariância é que restringindo f a X_0 podemos obter uma função de X_0 em si mesmo, em vez de simplesmente uma função de X_0 em X.

A.3 Relações de Equivalência

Uma relação de equivalência é um tipo particular de relação entre pares de elementos de um conjunto. Para definir uma relação de equivalência, precisamos primeiro decidir o que é uma 'relação'.

Certamente, uma definição formal de 'relação' deve envolver relações familiares tais como '$x = y$', '$x < y$', 'x é a mãe de y' e 'x é mais velho que y'. Se X é um conjunto, o que é necessário para determinar uma relação entre pares de elementos de X? O que se precisa, evidentemente, é de uma regra para determinar se, para dois quaisquer elementos dados x e y em X, x está na relação dada com y ou não. Uma tal regra R, será denominada uma **relação** (binária) sobre X. Se desejarmos ser ligeiramente mais precisos, poderemos proceder como segue. Indiquemos por $X \times X$ o conjunto dos pares ordenados (x, y) de elementos de X. Uma relação binária sobre X é uma função R de $X \times X$ no conjunto $\{0, 1\}$. Em outras palavras, R associa a cada par ordenado (x, y) um 1 ou um 0. A idéia é que se $R(x, y) = 1$, então x está na relação dada com y, e se $R(x, y) = 0$, não o está.

Se R é uma relação binária sobre o conjunto X, é conveniente escrever xRy quando $R(x, y) = 1$. Uma relação binária é dita

(1) **reflexiva**, se xRx para todo x em X;
(2) **simétrica**, se yRx sempre que xRy;
(3) **transitiva**, se xRz sempre que xRy e yRz.

Uma *relação de equivalência* sobre X é uma relação binária sobre X que é reflexiva, simétrica e transitiva.

Exemplo 5. (a) Sobre qualquer conjunto, a igualdade é uma relação de equivalência. Em outras palavras, se xRy significa $x = y$, então R é uma relação de equivalência. De fato, $x = x$, se $x = y$ então $y = x$, se $x = y$ e $y = z$ então $x = z$. A relação '$x \neq y$' é simétrica, mas não é reflexiva nem transitiva.

(b) Seja X o conjunto dos números reais e suponhamos que xRy signifique $x < y$. Então R não é uma relação de equivalência. Apesar de ser transitiva, não é nem reflexiva nem simétrica. A relação '$x \leq y$' é reflexiva e transitiva, mas não é simétrica.

(c) Seja E o plano euclidiano e seja X o conjunto dos triângulos no plano E. Então, a congruência é uma relação de equiva-

lência sobre X, isto é, '$T_1 \cong T_2$' (T_1 é congruente a T_2) é uma relação de equivalência sobre o conjunto dos triângulos no plano.

(d) Seja X o conjunto dos inteiros:

$$\ldots, -2, -1, 0, 1, 2, \ldots$$

Seja n um inteiro positivo fixo. Definamos uma relação R_n sobre X por: xR_ny se, e somente se, $(x - y)$ é divisível por n. A relação R_n é denominada **congruência módulo** n. Em vez de xR_ny escreve-se em geral,

$$x \equiv y, \bmod n (x \text{ é congruente a } y \text{ módulo } n)$$

quando $(x - y)$ é divisível por n. Para cada inteiro positivo n, a congruência módulo n é uma relação de equivalência sobre o conjunto dos inteiros.

(e) Sejam X e Y conjuntos e f uma função de X em Y. Definamos uma relação R sobre X por: x_1Rx_2 se, e somente se, $f(x_1) = f(x_2)$. É fácil verificar que R é uma relação de equivalência sobre o conjunto X. Como veremos, este exemplo engloba, na verdade, todas as relações de equivalência.

Suponhamos que R seja uma relação de equivalência sobre o conjunto X. Se x é um elemento de X, indiquemos por $E(x; R)$ o conjunto dos elementos y em X tais que xRy. Este conjunto $E(x; R)$ é denominado a **classe de equivalência** de x (segundo a relação de equivalência R). Como R é uma relação de equivalência, as classes de equivalência possuem as seguintes propriedades:

(1) Cada $E(x; R)$ é não-vazio, pois, como xRx, o elemento x pertence a $E(x; R)$.

(2) Sejam x e y elementos de X. Como R é simétrica, y pertence a $E(x; R)$ se, e somente se, x pertence a $E(y; R)$.

(3) Se x e y são elementos de X, as classes de equivalência $E(x; R)$ e $E(y; R)$ ou são idênticas ou não têm nenhum elemento em comum. Primeiro, suponhamos que xRy. Seja z um elemento arbitrário de $E(x; R)$, isto é, um elemento de X tal que xRz. Como R é simétrica, também temos zRx. Por hipótese, xRy e como R é transitiva, obtemos zRy, ou yRz. Isto mostra que todo elemento de $E(x; R)$ é um elemento de $E(y; R)$. Pela simetria de R, vemos, analogamente, que todo elemento de $E(y;R)$ é um elemento de $E(x; R)$; logo $E(x; R) = E(y; R)$. Afirmamos, agora, que se a re-

lação xRy não vale, então $E(x; R) \cap E(y; R)$ é vazia. De fato, se z está em ambas estas classes de equivalência, temos xRz e yRz, portanto, xRz e zRy, logo, xRy.

Indicando por $\mathscr{F}$ a família das classes de equivalência segundo uma relação de equivalência R, vemos que: (1) cada conjunto na família $\mathscr{F}$ é não-vazio. (2) cada elemento x de X pertence a um e somente um dos conjuntos na família $\mathscr{F}$; (3) xRy se, e somente se, x e v pertencem ao mesmo conjunto na família $\mathscr{F}$. Abreviadamente, a relação de equivalência R subdivide X na reunião de uma família de subconjuntos (não-vazios) disjuntos dois a dois. O argumento também vale no outro sentido. Suponhamos que $\mathscr{F}$ seja uma família arbitrária de subconjuntos de X que satisfaça as condições (1) e (2) acima. Se definirmos uma relação R por (3), então R será uma relação de equivalência sobre X e $\mathscr{F}$ a família das classes de equivalência segundo R.

Exemplo 6. Vejamos quais são as classes de equivalência segundo as relações de equivalência do Exemplo 5.

(a) Se R é a igualdade sobre o conjunto X, então a classe de equivalência do elemento x é simplesmente o conjunto $\{x\}$, cujo único elemento é x.

(b) Se X é o conjunto dos triângulos num plano e R é a relação de congruência, praticamente tudo o que se pode dizer, de início, é que a classe de equivalência do triângulo T consiste de todos os triângulos que são congruentes a T. Uma das tarefas da geometria plana é dar outras descrições destas classes de equivalência.

(c) Se X é o conjunto dos inteiros e R_n é a relação 'congruência módulo n', então existem precisamente n classes de equivalência. Cada inteiro x pode ser expresso de um único modo sob a forma $x = qn + r$, onde q e r são inteiros e $0 \leq r \leq n - 1$. Isto mostra que cada x é congruente módulo n a exatamente um dos n inteiros $0, 1, 2, \ldots, n-1$. As classes de equivalência são

$$E_0 = \{\ldots, -2n, -n, 0, n\ 2n, \ldots\}$$

$$E_1 = \{\ldots, 1 - 2n, 1 - n, 1, 1 + n, 1 + 2n, \ldots\}$$

$$\vdots \qquad \vdots$$

$$E_{n-1} = \{\ldots, n - 1 - 2n, n - 1 - n; n - 1, n - 1 + n, n - 1 + 2n, \ldots\}.$$

(d) Suponhamos que X e Y sejam conjuntos, f uma função de X em Y e R a relação de equivalência definida por: $x_1 R x_2$ se, e somente se, $f(x_1) = f(x_2)$. As classes de equivalência para R são exatamente os maiores subconjuntos de X sobre os quais f é 'constante'. Outra descrição das classes de equivalência é a que segue. Elas estão em correspondência bijetora com os elementos na imagem de f. Se y está na imagem de f, o conjunto dos x em X tais que $f(x) = y$ é uma classe de equivalência para R. isto define uma correspondência bijetora entre os elementos da imagem de f e as classes de equivalência R.

Façamos mais um comentário acerca de relações de equivalência. Dada uma relação de equivalência R sobre X, seja $\mathscr{F}$ a família das classes de equivalência segundo R. Associando ao elemento x a classe de equivalência $E(x; R)$, definimos uma função f de X em $\mathscr{F}$ (na verdade, sobre $\mathscr{F}$):

$$f(x) = E(x; R).$$

Isto mostra que R é a relação de equivalência associada a uma função cujo domínio é X, como no Exemplo 5 (e). O que isto nos diz é que toda relação de equivalência sobre o conjunto X é determinada como segue. Temos uma regra (função) f que associa a cada elemento x de X um objeto $f(x)$, e xRy se, e somente se, $f(x) = f(y)$. Agora, deve-se considerar $f(x)$ como uma propriedade de x, de modo que o que a relação de equivalência faz (a grosso modo) é reunir todos os elementos de X que têm esta propriedade em comum. Se o objeto $f(x)$ é a classe de equivalência de x, então tudo o que se disse é que a propriedade comum dos elementos de uma classe de equivalência é que eles pertencem à mesma classe de equivalência. É claro que isto não diz muito. Em geral, existem muitas funções distintas f que determinam uma dada relação de equivalência como acima, e um objetivo no estudo das relações de equivalência é determinar uma tal f que dê uma descrição significativa e elementar da relação de equivalência. Na Seção A.5 veremos como isto é conseguido para algumas relações de equivalência particulares que surgem em álgebra linear.

A.4 Espaços Quocientes

Seja V um espaço vetorial sobre o corpo F e seja W um subespaço de V. Existem, em geral, muitos subespaços W' que são suple-

lação xRy não vale, então $E(x; R) \cap E(y; R)$ é vazia. De fato, se z está em ambas estas classes de equivalência, temos xRz e yRz, portanto, xRz e zRy, logo, xRy.

Indicando por $\mathscr{F}$ a família das classes de equivalência segundo uma relação de equivalência R, vemos que: (1) cada conjunto na família $\mathscr{F}$ é não-vazio, (2) cada elemento x de X pertence a um e somente um dos conjuntos na família $\mathscr{F}$; (3) xRy se, e somente se, x e v pertencem ao mesmo conjunto na família $\mathscr{F}$. Abreviadamente, a relação de equivalência R subdivide X na reunião de uma família de subconjuntos (não-vazios) disjuntos dois a dois. O argumento também vale no outro sentido. Suponhamos que $\mathscr{F}$ seja uma família arbitrária de subconjuntos de X que satisfaça as condições (1) e (2) acima. Se definirmos uma relação R por (3), então R será uma relação de equivalência sobre X e $\mathscr{F}$ a família das classes de equivalência segundo R.

Exemplo 6. Vejamos quais são as classes de equivalência segundo as relações de equivalência do Exemplo 5.

(a) Se R é a igualdade sobre o conjunto X, então a classe de equivalência do elemento x é simplesmente o conjunto $\{x\}$, cujo único elemento é x.

(b) Se X é o conjunto dos triângulos num plano e R é a relação de congruência, praticamente tudo o que se pode dizer, de início, é que a classe de equivalência do triângulo T consiste de todos os triângulos que são congruentes a T. Uma das tarefas da geometria plana é dar outras descrições destas classes de equivalência.

(c) Se X é o conjunto dos inteiros e R_n é a relação 'congruência módulo n', então existem precisamente n classes de equivalência. Cada inteiro x pode ser expresso de um único modo sob a forma $x = qn + r$, onde q e r são inteiros e $0 \leq r \leq n - 1$. Isto mostra que cada x é congruente módulo n a exatamente um dos n inteiros $0, 1, 2, \ldots, n - 1$. As classes de equivalência são

$$E_0 = \{\ldots, -2n, -n, 0, n\ 2n, \ldots\}$$

$$E_1 = \{\ldots, 1 - 2n, 1 - n, 1, 1 + n, 1 + 2n, \ldots\}$$

$$\vdots \qquad \vdots$$

$$E_{n-1} = \{\ldots, n - 1 - 2n, n - 1 - n; n - 1, n - 1 + n, n - 1 + 2n, \ldots\}.$$

(d) Suponhamos que X e Y sejam conjuntos, f uma função de X em Y e R a relação de equivalência definida por: $x_1 R x_2$ se, e somente se, $f(x_1) = f(x_2)$. As classes de equivalência para R são exatamente os maiores subconjuntos de X sobre os quais f é 'constante'. Outra descrição das classes de equivalência é a que segue. Elas estão em correspondência bijetora com os elementos na imagem de f. Se y está na imagem de f, o conjunto dos x em X tais que $f(x) = y$ é uma classe de equivalência para R. isto define uma correspondência bijetora entre os elementos da imagem de f e as classes de equivalência R.

Façamos mais um comentário acerca de relações de equivalência. Dada uma relação de equivalência R sobre X, seja $\mathscr{F}$ a família das classes de equivalência segundo R. Associando ao elemento x a classe de equivalência $E(x; R)$, definimos uma função f de X em $\mathscr{F}$ (na verdade, sobre $\mathscr{F}$):

$$f(x) = E(x; R).$$

Isto mostra que R é a relação de equivalência associada a uma função cujo domínio é X, como no Exemplo 5 (e). O que isto nos diz é que toda relação de equivalência sobre o conjunto X é determinada como segue. Temos uma regra (função) f que associa a cada elemento x de X um objeto $f(x)$, e xRy se, e somente se, $f(x) = f(y)$. Agora, deve-se considerar $f(x)$ como uma propriedade de x, de modo que o que a relação de equivalência faz (a grosso modo) é reunir todos os elementos de X que têm esta propriedade em comum. Se o objeto $f(x)$ é a classe de equivalência de x, então tudo o que se disse é que a propriedade comum dos elementos de uma classe de equivalência é que eles pertencem à mesma classe de equivalência. É claro que isto não diz muito. Em geral, existem muitas funções distintas f que determinam uma dada relação de equivalência como acima, e um objetivo no estudo das relações de equivalência é determinar uma tal f que dê uma descrição significativa e elementar da relação de equivalência. Na Seção A.5 veremos como isto é conseguido para algumas relações de equivalência particulares que surgem em álgebra linear.

A.4 Espaços Quocientes

Seja V um espaço vetorial sobre o corpo F e seja W um subespaço de V. Existem, em geral, muitos subespaços W' que são suple-

mentares de W, isto é, subespaços com a propriedade de que $V = W \oplus W'$. Se temos um produto interno sobre V e se W é de dimensão finita, existe um subespaço particular que provavelmente se denominaria o subespaço suplementar 'natural' de W. Ele é o suplementar ortogonal de W. Mas se V não possui nenhuma estrutura além de sua estrutura de espaço vetorial, não existe nenhuma maneira de selecionar um subespaço W' que se pudesse denominar o subespaço suplementar natural de W. Contudo, pode-se construir, a partir de V e W, um espaço vetorial V/W, conhecido como o 'quociente' de V por W, que desempenhará o papel do suplementar natural de W. Este espaço quociente não é um subespaço de V, portanto, não pode ser realmente um subespaço suplementar de W, no entanto, ele é um espaço vetorial definido apenas em termos de V e W que tem a propriedade de ser isomorfo a todo subespaço W' que seja suplementar de W.

Seja W um subespaço do espaço vetorial V. Se α e β são vetores em V, dizemos que α é **congruente** a β **módulo** W se o vetor $(\alpha - \beta)$ está no subespaço W. Se α é congruente a β módulo W, escrevemos

$$\alpha \equiv \beta, \text{ mod } W.$$

Ora, a congruência módulo W é uma relação de equivalência sobre V.

(1) $\alpha \equiv \alpha$, mod W, porque $\alpha - \alpha = 0$ está em W.

(2) Se $\alpha \equiv \beta$, mod W, então $\beta \equiv \alpha$, mod W. De fato, como W é um subespaço de V, o vetor $(\alpha - \beta)$ está em W se, e somente se, $(\beta - \alpha)$ está em W.

(3) Se $\alpha \equiv \beta$, mod W, e $\beta \equiv \gamma$, mod W, então $\alpha \equiv \gamma$, mod W. De fato, se $(\alpha - \beta)$ e $(\beta - \gamma)$ estão em W, então $\alpha - \gamma = (\alpha - \beta) + (\beta - \gamma)$ está em W.

As classes de equivalência desta relação de equivalência são conhecidas como as **classes laterais** de W. Qual é a classe de equivalência (classe lateral) de um vetor α? Ela consiste dos vetores β em V tais que $(\beta - \alpha)$ está em W, isto é, os vetores da forma $\beta = \alpha + \gamma$, com γ em W. Por esta razão, a classe lateral do vetor α é indicada por

$$\alpha + W.$$

É conveniente pensar na classe lateral de α em relação a W como sendo o conjunto dos vetores obtidos por translação do subespaço

W pelo vetor α. Para visualizar estas classes laterais, o leitor pode pensar no seguinte caso particular: Seja V o espaço R^2 e seja W um subespaço unidimensional de V. Se imaginarmos V como sendo o plano euclidiano, W será uma reta passando pela origem. Se Se $\alpha = (x_1, x_2)$ é um vetor em V, a classe lateral $\alpha + W$ é a reta que passa pelo ponto (x_1, x_2) e é paralela a W.

A coleção de todas as classes laterais de W será indicada por V/W. Definamos, agora, uma adição de vetores e uma multiplicação escalar sobre V/W como segue:

$$\begin{aligned}(\alpha + W) + (\beta + W) &= (\alpha + \beta) + W\\ c(\alpha + W) &= (c\alpha) + W.\end{aligned}$$

Em outras palavras, a soma da classe lateral de α com a classe lateral de β é a classe lateral de $(\alpha + \beta)$, e o produto do escalar c pela classe lateral de α é a classe lateral do vetor $c\alpha$. Ora, muitos vetores distintos em V terão a mesma classe lateral em relação a W, portanto, precisamos verificar que a soma e o produto acima dependem somente das classes laterais envolvidas. O que isto significa é que precisamos mostrar o seguinte:

(1) Se $\alpha \equiv \alpha'$, mod W, e $\beta \equiv \beta'$, mod W, então

$$\alpha + \beta \equiv \alpha' + \beta', \text{ mod } W.$$

(2) Se $\alpha \equiv \alpha'$, mod W, então $c\alpha \equiv c\alpha'$, mod W.

Estes fatos são fáceis de verificar. (1) Se $\alpha - \alpha'$ está em W e $\beta - \beta'$ está em W, então como $(\alpha + \beta) - (\alpha' + \beta') = (\alpha - \alpha') + (\beta - \beta')$ vemos que $\alpha + \beta$ é congruente a $\alpha' + \beta'$ módulo W. (2) Se $\alpha - \alpha'$ está em W e c é um escalar arbitrário, então $c\alpha - c\alpha' = c(\alpha - \alpha')$ está em W.

Agora é fácil verificar que V/W, com a adição de vetores e a multiplicação escalar acima definidas, é um espaço vetorial sobre o corpo F. Deve-se verificar diretamente cada um dos axiomas para um espaço vetorial. Cada uma das propriedades da adição de vetores e da multiplicação escalar decorre da propriedade correspondente das operações em V. Um comentário deve ser feito. O vetor nulo em V/W será a classe lateral do vetor nulo em V. Em outras palavras, W é o vetor nulo em V/W.

O espaço vetorial V/W é denominado o **quociente** (ou diferença) de V por W. Existe uma transformação linear natural Q

de V sobre V/W. É definida por $Q(\alpha) = \alpha + W$. Deve-se ver que definimos as operações em V/W exatamente de modo que esta transformação Q viesse a ser linear. Notemos que o núcleo de Q é exatamente o subespaço W. Denominamos Q a **transformação quociente** (ou **aplicação quociente**) de V sobre V/W.

A relação entre o espaço quociente V/W e subespaços de V que são suplementares de W pode, agora, ser enunciada como segue.

Teorema. *Seja* W *um subespaço do espaço vetorial* V *e seja* Q *a aplicação quociente de* V *sobre* V/W. *Suponhamos que* W′ *seja um subespaço de* V. *Então* $V = W \oplus W'$ *se, e somente se, a restrição de* Q *a* W′ *é um isomorfismo de* W′ *em* V/W.

Demonstração. Suponhamos que $V = W \oplus W'$. Isto significa que cada vetor α em V pode ser expresso de um único modo sob a forma $\alpha = \gamma + \gamma'$, com γ em W e γ' em W'. Então $Q\alpha = Q\gamma + Q\gamma' = Q\gamma'$, isto é, $\alpha + W = \gamma' + W$. Isto mostra que Q leva W' sobre V/W, ou seja, que $Q(W') = V/W$. Além disso, Q é injetora em W'; de fato, suponhamos que γ_1' e γ_2' sejam vetores em W' e que $Q\gamma_1' = Q\gamma_2'$. Então, $Q(\gamma_1' - \gamma_2') = 0$ de modo que $\gamma_1' - \gamma_2'$ está em W. Este vetor também está em W', que é disjunto de W; logo $\gamma_1' - \gamma_2' = 0$. A restrição de Q a W' é, portanto, uma transformação linear bijetora de W' em V/W.

Suponhamos que W' seja um subespaço de V tal que Q seja injetora em W' e que $Q(W') = V/W$. Seja α um vetor em V. Então existe um vetor γ' em W' tal que $Q\gamma' = Q\alpha$, isto é, $\gamma' + W = \alpha + W$. Isto significa que $\alpha = \gamma + \gamma'$ para algum vetor γ em W. Portanto, $V = W + W'$. Para ver que W e W' são disjuntos, suponhamos que γ esteja em W e em W'. Como γ está em W, temos $Q\gamma = 0$. Mas Q é injetora em W', logo devemos ter que $\gamma = 0$. Assim, temos que $V = W \oplus W'$.

O que este teorema realmente diz é que W' é um suplementar de W se, e somente se, W' é um subespaço que contém exatamente um elemento de cada classe lateral de W. Ele mostra que, quando $V = W \oplus W'$, a aplicação quociente Q 'identifica' W' com V/W. Abreviadamente, $(W \oplus W')/W$ é isomorfo a W' de uma maneira 'natural'.

Um fato bastante óbvio deve ser notado. Se W é um subespaço do espaço vetorial V de dimensão finita, então

$$\dim W + \dim(V/W) = \dim V.$$

Pode-se ver isto a partir do teorema acima. Talvez seja mais fácil observar que esta fórmula sobre dimensões diz:

$$\text{nulidade}\,(Q) + \text{posto}\,(Q) = \dim V.$$

Não é nosso objetivo, aqui, fazer um tratamento detalhado dos espaços quocientes. Contudo, existe um resultado fundamental que devemos demonstrar.

Teorema. *Sejam* V *e* Z *espaços vetoriais sobre o corpo* F. *Suponhamos que* T *seja uma transformação linear de* V *sobre* **Z**. *Se* W *é o núcleo de* T, *então* **Z** *é isomorfo a* **V/W**.

Demonstração. Definamos uma transformação U de V/W em Z por $U(\alpha + W) = T\alpha$. Precisamos verificar que U está bem definida, isto é, que se $\alpha + W = \beta + W$ então $T\alpha = T\beta$. Isto decorre do fato de que W é o núcleo de T; de fato, $\alpha + W = \beta + W$ significa que $\alpha - \beta$ está em W e isto ocorre se, e somente se, $T(\alpha - \beta) = 0$. Isto mostra não só que U está bem definida, mas também que U é injetora.

Agora é fácil verificar que U é linear e leva V/W sobre Z, pois T é uma transformação linear de V sobre Z.

A.5 Relações de Equivalência em Álgebra Linear

Vamos considerar algumas das relações de equivalência que surgem no texto deste livro. Esta é, apenas, uma amostra de tais relações.

(1) Sejam m e n inteiros positivos e F um corpo. Seja X o conjunto das $m \times n$ matrizes sobre F. Então, a linha-equivalência é uma relação de equivalência sobre o conjunto X. A afirmação 'A é linha-equivalente a B' significa que A pode ser obtida de B por uma sucessão finita de operações elementares sobre linhas. Se escrevermos $A \sim B$ para indicar que A é linha-equivalente a B, então não é difícil verificar as propriedades (i) $A \sim A$; (ii) se $A \sim B$, então $B \sim A$; (iii) se $A \sim B$ e $B \sim C$, então $A \sim C$. Que sabemos a respeito desta relação de equivalência? Na realidade, sabemos bastante. Por exemplo, sabemos que $A \sim B$ se, e somente se, $A = PB$ para alguma $m \times m$ matriz inversível P, ou, $A \sim B$ se, e somente se, os sistemas homogêneos de equações lineares $AX = 0$ e $BX = 0$ têm as mesmas soluções. Também temos informações bem explícitas sobre as classes de equivalência, segundo esta relação. Cada $m \times n$ matriz A

é linha-equivalente a uma, e somente uma, matriz linha-reduzida à forma em escada. O que isto diz é que cada classe de equivalência segundo esta relação contém precisamente uma matriz R linha-reduzida à forma em escada; a classe de equivalência determinada por R consiste das matrizes $A = PR$ onde P é uma $m \times m$ matriz inversível. Pode-se também considerar esta descrição das classes de equivalência da seguinte maneira: Dada uma $m \times n$ matriz A, temos uma regra (função) f que associa a A a matriz $f(A)$, linha-reduzida à forma em escada, que é linha-equivalente a A. A linha-equivalência é completamente determinada por f. De fato, $A \sim B$ se, e somente se, $f(A) = f(B)$, isto é, se, e somente se, A e B têm a mesma forma em escada linha-reduzida.

(2) Seja n um inteiro positivo e F um corpo. Seja X o conjunto das $n \times n$ matrizes sobre F. Então, a semelhança é uma relação de equivalência sobre X; cada $n \times n$ matriz A é semelhante a si mesma; se A é semelhante a B, então B é semelhante a A; se A é semelhante a B e B é semelhante a C, então A é semelhante a C. Também sabemos muita coisa acerca desta relação de equivalência. Por exemplo, A é semelhante a B se, e somente se, A e B representam o mesmo operador linear sobre F^n em relação a bases ordenadas (possivelmente) diferentes. Entretanto, sabemos algo muito mais profundo que isto. Cada $n \times n$ matriz A sobre F é semelhante (sobre F) a uma, e somente uma, matriz sob a forma racional (Capítulo 7). Em outras palavras, cada classe de equivalência segundo a relação de semelhança contém exatamente uma matriz que está sob a forma racional. Uma matriz sob a forma racional é determinada por uma k-upla $(p_1, \ldots, p_k)$ de polinômios unitários com a propriedade de que p_{j+1} divide p_j, $j = 1, \ldots, k-1$. Assim, temos uma função f que associa a cada $n \times n$ matriz A uma k-upla $f(A) = (p_1, \ldots, p_k)$ que satisfaz a condição de divisibilidade p_{j+1} divide p_j. Então, A e B são semelhantes se, e somente se, $f(A) = f(B)$.

(3) Eis um caso particular do Exemplo 2 acima. Seja X o conjunto das 3×3 matrizes sobre um corpo F. Consideremos a relação de semelhança sobre X. Se A e B são 3×3 matrizes sobre F, então A e B são semelhantes se, e somente se, possuem o mesmo polinômio característico e o mesmo polinômio minimal. Associado a cada 3×3 matriz A, temos um par (f, p) de polinômios unitários que satisfazem

(a) $\mathrm{gr}(f) = 3$.

(b) p divide f,

sendo f o polinômio característico de A e p o polinômio minimal de A. Dados polinômios unitários f e p sobre F que satisfaçam (a) e (b), é fácil exibir uma 3×3 matriz sobre F, cujos polinômios característicos e minimal sejam f e p, respectivamente. O que tudo isso nos diz é o que segue. Se considerarmos a relação de semelhança sobre o conjunto das 3×3 matrizes sobre F, as classes de equivalência estarão em correspondência bijetora com os pares ordenados (f, p) de polinômios unitários sobre F que satisfazem (a) e (b).

A.6 O Axioma da Escolha

Sem muito rigor, o Axioma da Escolha é uma regra (ou um princípio) de pensamento afirmando que dada uma família de conjuntos não vazios, podemos escolher um elemento de cada conjunto. Para sermos mais precisos, suponhamos que A seja um conjunto de índices e que exista associado a cada α de A, um conjunto não-vazio S_α. "Escolher" um elemento de cada S_α significa dar uma regra f que associe a cada α algum elemento $f(\alpha)$ do conjunto S_α. O axioma da escolha afirma que isso pode ser feito, isto é, dada uma família de conjuntos $\{S_\alpha\}$, existe uma função f de A em

$$\bigcup_\alpha S_\alpha$$

tal que $f(\alpha)$ esteja em S_α para cada α. Esse princípio é aceito pela maioria dos matemáticos, apesar de surgirem muitas situações onde não é nada evidente como qualquer função explícita f possa ser encontrada.

O Axioma da Escolha tem algumas conseqüências surpreendentes. A maioria delas não tem relação com o assunto tratado nesse livro; no entanto, uma conseqüência merece ser mencionada: Todo espaço vetorial possui uma base. Por exemplo, o corpo dos números reais possui uma base como espaço vetorial sobre o corpo dos números racionais. Em outras palavras, existe um subconjunto S de R, que é linearmente independente sobre o corpo dos

números racionais, e que possui a propriedade de que todo número real é uma combinação linear racional de algum número finito de elementos de *S*. Não deduziremos aqui esse resultado para espaços vetoriais, a partir do Axioma da Escolha. Para uma demonstração, referimos o leitor ao livro de Kelley, mencionado na bibliografia.

BIBLIOGRAFIA

Halmos, P., *Finite-Dimensional Vector Spaces*. D. Van Nostrand Co., Princeton, 1958.

Jacobson, N., *Lectures in Abstract Algebra*, II. D. Van Nostrand Co., Princeton, 1953.

Kelley, John L., *General Topology*, D. Van Nostrand Co., Princeton, 1955.

MacLane, S. and Birkhoff, G., *Algebra*, The Macmillan Co., New York, 1967.

Schreier, O. and Sperner, E., *Modern Algebra and Matrix Theory*. 2.ª Ed., Chelsea Publishing Co., New York, 1955.

van der Waerden, B. L., *Modern Algebra* (dois volumes), Rev. Ed., Frederick Ungar Publishing Co., New York, 1969.

ÍNDICE ALFABÉTICO

A

Adjunta:
    clássica, 189, 202
    de uma transformação, 378
Álgebra, 148, 150, 151
    auto-adjunta, 440
    comutativa, 148
    das séries formais de potência, 150
    linear, 148, 506
Anéis comutativos, 178
Anel, 178
    comutativo, 180
    de Grassman, 221, 231
Anulador:
    de subconjunto, 129
    de subespaço, 129
    de vetor (T-anulador), 258, 291
Aproximação, 362
Assinatura de forma bilinear, 474
Associatividade, 1
    da adição de vetores, 35
    da multiplicação de matrizes, 36
Autovalor (*ver*: Valor característico)
Axioma da escolha, 508

B

Base, 50
    canônica de F^n, 52
    de espaço vetorial, 51
    de módulo, 209
    dual, 126, 210
    ordenada, 52, 62, 409
    ortonormal, 359, 398

C

Característica do corpo, 3-4
Classe de equivalência, 499
Classe lateral, 226, 502
Coeficiente de um polinômio, 151
Cofator, 201
Coluna-equivalência, 328
Combinação linear, 39
    de equações, 4
    de vetores, 39
Componentes primárias, 448
Composição, 495
Condutor, 258, 259, 297
Congruência, 176, 499, 502
Conjugação, 353 (exerc. 13)
Conjugado, 346
Conjunto, 492
    ortogonal, 356
    ortonormal, 357
    vazio, 492
Coordenadas, 61
Corpo, 2
    algebricamente fechado, 175

D

Decomposição:
    cíclica de operador linear, 295
    em fatores primos, de polinômio, 170
    em somas diretas, 267
    polar de operador linear, 437
    primária de polinômio, 170
Dependência linear, 50, 59
Derivada do polinômio, 163, 340

Desenvolvimento segundo Laplace, 229
Desigualdade:
de Bessel, 368
de Cauchy-Schwarz, 355
Determinante, 183
de transformação linear, 216
Diagonalização, 264, 278
de forma bilinear simétrica, 472
de forma hermitiana, 411
de matriz (operador) auto-adjunta(o), 401
de matriz (operador) normal, 404
simultânea, 264, 437
unitária, 404
Dimensão, 50
de espaço vetorial, 50
finita, 51
Distância, 369 (Ex. 15 – exerc. 4)
Divisão com resto, 167

E

Elemento:
de conjunto, 492
de matriz, 7
unidade, 148, 178
Equações:
diferenciais, 286 (Ex. 14), 318 (Ex. 8)
lineares (*ver*: Sistema de equações lineares)
Equivalência:
de matrizes ortogonais, 394
unitária de matrizes, 394
de transformações lineares, 454
Escalar, 2
Espaço, 40
característico, 233
com produto interno, 345
dual, 125
euclidiano, 354
linha, 48
quociente, 501
solução, 44
unitário, 354
vetorial, 35
de dimensão finita, 51
de funções polinomiais, 38
de n-uplas, 36
de soluções de equações lineares, 44
quociente, 501
sobre o corpo, 36
Espectro, 430

F

$F^{m \times n}$, 37
F^n, 36
Fatoração de polinômio, 170
Fatores invariantes, 305, 321, 334
de uma matriz, 305, 321
Forma:
alternada, 215
bilinear, 211, 407, 457
anti-simétrica, 478
não-degenerada (não-singular), 465
positiva definida, 469
simétrica, 468
de Jordan de uma matriz, 290, 311, 314, 316
hermitiana, 411
multilinear, 211
não-degenerada, 413 (exerc. 6)
não-negativa, 413
não-singular (*ver*: Forma não-degenerada)
normal, 330, 332
positiva, 413, 417
quadrática, 349, 468
r-linear, 211
racional, 290
racional de matriz, 304
sesquilinear, 407
Formas canônicas elementares, 232
Fórmula:
da dimensão, 55-56
de interpolação de Lagrange, 158
de Taylor, 164, 340
Função, 493
determinante, 179-180, 183
idêntica, 495
inversível, 495
linear, 86, 123, 371,
multilinear, 210-211
n-linear, 180
alternada, 183, 186
polinomial, 38

Funcional linear, 123

G

Grau, 118
de forma multilinear, 211
de polinômio, 151
Grupo, 104
comutativo, 105
de Lorentz, 486
linear geral, 392
ortogonal, 485
pseudo-ortogonal, 485
que conserva formas bilineares, 483
simétrico, 195

H

Hiperplano, 129, 139

I

Ideal, 166
principal, 166
Identidades de polarização, 350, 468
Imagem, 90
de função, 494
de transformação linear, 90
Independência linear, 50, 59
Inteiros, 3
positivos, 3
Interpolação, 157
Interseção, 492
de subespaços, 43
Inversa:
de função, 495
de matriz, 27, 203
Inverso:
à direita, 27
à esquerda, 27
Isomorfismo, 107
de espaços com produto interno, 384
de espaços vetoriais, 107

L

Linearmente:
dependente, 50, 59
independente, 50, 59
Linha-equivalência, 9
Linha-equivalente, 8, 70, 323

M

Matriz(es), 7
anti-simétrica, 206 (Ex. 8), 269
associada, 293
auto-adjunta, 44, 401
completa, 17
das coordenadas, 64
de forma, 411
de forma bilinear, 463
de transformação linear, 112, 113
em relação a uma base ortonormal, 376-377
de Vandermonde, 158
do produto interno, 351
dos coeficientes, 7
dos cofatores, 201
elementar, 25, 323
de Jordan, 314
hermitiana (*ver*: Matriz auto-adjunta)
inversível, 27, 203
linha-reduzida, 11
à forma em escada, 14, 70
nilpotente, 312
normal, 403
nula, 13
ortogonal, 207 (Ex. 8), 485
positiva, 419
semelhantes, 120
simétrica, 44, 269
triangular, 197 (Ex. 5)
superior, 34
unidade, 11
unitária, 207 (Ex. 8), 389
Máximo divisor comum, 169
Menores principais de matriz, 415
Módulo, 208
dual, 210
finitamente gerado, 209
livre, 209
Movimento rígido, 396 (exerc. 14)
Mudança de base, 115
Multiplicidade, 165

N

n-upla, 36
Norma, 349

Núcleo, 90
Nulidade de uma transformação linear, 90
Números:
complexos, 2, 38
racionais, 3
reais, 2

O

Operações elementares:
sobre colunas, 32, 328
sobre linhas, 7, 327
Operador linear, 97 (*ver também*: Transformação linear)
auto-adjunto, 382, 389
diagonalizável, 236
hermitiano (*ver*: Operador auto-adjunto)
não-negativo, 419, 435
nilpotente, 284
normal, 398
positivo, 419
semi-simples, 335
unitário, 388
Ortogonalização, 358

P

Parte diagonalizável de operador linear, 283
Permutação, 192
de grau, 192
par, ímpar, 193
Polinômio, 148, 151
característico, 234
constante, 151
irredutível (primo), 171
minimal, 245
de matriz, 245
de operador linear, 245
primo, 171
redutível, 171
unitário, 151
Polinômios anuladores, 244
Posto, 90, 145
coluna, 91, 145
de determinante, 207 (Ex. 8)
de forma bilinear, 465
de matriz, 147
de módulo, 209
de transformação linear, 91
linha, 70, 91, 145
Processo de Gram-Schmidt, 358, 369
Produto:
de matrizes, 21, 113
de permutações, 194
de transformações lineares, 98
exterior, 223, 226
interno, 345, 384, 406
canônico, 346, 347
tensorial, 213
Projeção, 269
ortogonal, 365, 425

R

Raíz:
de família de operadores, 437
de polinômio, 163
quadrada, 436
Regra:
de Cramer, 205
do paralelogramo, 352 (Ex. 6 – exerc. 9)
Relação, 498
de equivalência, 498, 506
Relativamente primos, 169
Resolução:
do operador idêntico, 429, 435
espectral, 428, 435
Restrição, 496
de função, 496
de operador, 256
Reunião, 492
Rotação, 68, 395 (Ex. 28 – exerc. 4)

S

Semelhança de matrizes, 120
Seqüência de vetores, 56-57
Série formal de potência, 150-151
Símbolo de Kronecker, 11
Sinal de permutação, 193
Sistemas de equações lineares, 4
homogêneas, 4
Sistemas equivalentes de equações, 5

Sobrejetora, 494
Soma, 46
de subespaços, 43
direta, 269, 273
das matrizes, 274
dos operadores, 273
invariante, 273
Subconjunto, 492
invariante, 497
próprio, 492
Subcorpo, 2
Subespaço(s), 43, 254
admissível, 296
anuladores, 290
cíclico, 290
disjuntos (*ver*: Subespaços independentes)
gerado por, 45
independentes, 267
invariante, 254, 264, 401
nulo, 44
quociente, 501
suplementar, 295
ortogonal, 364
T-admissível, 296
Submatriz, 207 (Ex. 8)
Suplementar ortogonal de subespaço, 364

T

T-admissível, 296
T-anulador, 258, 291
T-condutor, 258, 259, 299
Tensor, 211
Teorema:
da decomposição cíclica, 297
da decomposição primária, 280, 281
de Cayley-Hamilton, 251, 294, 303
do eixo principal, 412
espectral, 427
fundamental da álgebra, 175
generalizado, 303
Teoria espectral, 427
Traço:
de matriz, 124
de operador linear, 135
de transformação linear, 134
Transformação:
de Lorentz, 397 (exerc. 15), 486
derivação, 85
idempotente (*ver*: projeção)
linear (*ver também*: Operador linear), 85, 96-97, 114
auto-adjunta, 382, 401
diagonalizável, 236
inversível, 100
não-negativa, 419, 435
não-singular, 101
nilpotente, 284
normal, 398
ortogonal, 390
positiva, 419
quociente, 501
semi-simples, 337
triangulável, 259, 403
unitária, 388, 454
nula, 85
Transformada de Cayley, 395 (exerc. 7)
Transposta:
conjugada, 347
de matriz, 147
de transformação linear, 142
Triangulação, 264, 425
simultânea, 264

U

Unicidade dos determinantes, 190

V

Valor característico, 233, 234
Vetor(es), 36
característico, 233
cíclico, 290
linha, 48
ortogonais, 356, 468
separador, 310 (Ex. 4 – exerc. 14)

Z

Zero de polinômio, 163

Composto em Univers pelo processo de Composição Eletrônica – IBM no Setor de Composição de LIVROS TÉCNICOS E CIENTÍFICOS EDITORA S.A. – Rio de Janeiro – Brasil

impresso na
GRÁFICA PORTINHO CAVALCANTI LTDA.
Rua Irineu Marinho, 30 - s/loja 206
Tel.: 224-7732 (PABX)
Rio de Janeiro – RJ